Vol. 40 OCTOBER, 1949 No. 6.

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

The Influence of Training in Perceptual Span and Perceptual Speed upon Reading Ability 321

CECIL MAX FREEBURNE

General, Verbal, and Spatial Bi-factors after Three Years 353

FRANCES SWINEFORD

The Function of the Division of Educational Psychology of the American Psychological Association: A Committee Report . . 361

VICTOR H. NOLL, JOHN E. HORROCKS, AND G. LESTER ANDERSON

An Application of Chi-square to the Determination of the Discriminating Power of Test Questions 371

ORRIS C. HERFINDAHL

The Constancy of Item-test Correlation Coefficients Computed from Upper and Lower Groups 378

JEROME E. DOPPELT AND EDITH M. POTTS

Book Review 382

$6.00 per Year • Published Monthly October to May

WARWICK & YORK, INC.

BALTIMORE 2, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

IVERSITY OF MICHIGAN
THE HORACE H. RACKHAM
CHOOL OF GRADUATE STUDIES
ANN ARBOR
NSTITUTE FOR HUMAN ADJUSTMENT
SPEECH CLINIC
1007 EAST HURON STREET

THE JOURNAL OF Educational Psychology

Established 1910

EDITORIAL BOARD

STEPHEN M. COREY
Teachers College, Columbia University
—*Learning*—

ROBERT T. ROCK, JR.
Fordham University
—*Individual Differences*—

JACK W. DUNLAP
10 E. 49th St., New York 17, N.Y.
—*Technical, Statistics*—

PERCIVAL M. SYMONDS
Teachers College, Columbia University
—*Mental Hygiene*—

KARL J. HOLZINGER
University of Chicago
—*Factor Analysis*—

MILES A. TINKER
University of Minnesota
—*Reading*—

HAROLD E. JONES
University of California
—*Social Behavior, Child Psychology*—

ALEXANDER G. WESMAN
The Psychological Corporation
—*Tests*—

H. H. REMMERS
Purdue University
—*Attitudes, Teacher Evaluation*—

PAUL A. WITTY
Northwestern University
—*Children's Interests*—

H. E. BUCHHOLZ
Managing Editor
10 E. Centre St., Baltimore 2, Md.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY is devoted primarily to the scientific study of problems of learning, teaching, and measurement of the psychological development of the individual. The JOURNAL will contain articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, and personality, particularly as related to school adjustment; emotion, motivation, and character; mental development and methods. This last will include tests, statistical techniques, and research techniques in cross-sectional and developmental studies.

Manuscripts may be submitted to any member of the Editorial Board, but the handling of an article will be facilitated if it is sent to that member of the Board who is designated as particularly interested in the phase of psychology dealt with. (Such designations appear after the Editors' names in the list given above.)

Books and other materials for review and correspondence regarding editorial matters should be addressed to The Journal of Educational Psychology, Warwick and York, Inc., Publishers, 10 E. Centre St., Baltimore 2, Md.

Manuscripts should be typed and double-spaced throughout, including quotations, footnotes, and references. In order to attain

(Continued on Inside Back Cover)

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 40 October, 1949 Number 6

THE INFLUENCE OF TRAINING IN PERCEPTUAL SPAN AND PERCEPTUAL SPEED UPON READING ABILITY*

CECIL MAX FREEBURNE†

From the Psychological Laboratory of the State University of Iowa

INTRODUCTION

This study is concerned primarily with the relationships among perceptual span, perceptual speed, and reading ability, and with the effect upon reading performance of training in perceptual span and perceptual speed. Perceptual span and perceptual speed are defined in terms of Flashmeter projection (see Methods, Procedure, Apparatus and Materials). Training in perceptual span was designed to increase the amount of material the subjects could grasp during a single fixation by systematically increasing the length of phrases flashed upon a screen at constant exposure times during the training period. Training in perceptual speed was designed to increase the speed with which the subjects could grasp phrases of a constant length, by systematically shortening the exposure times for the phrases during the training period. Control groups were used to help in evaluating the performance of the experimental groups. Secondary questions raised in this study concerned: (1) the prediction of reading improvement from perceptual data; (2) the relationship between gains in perceptual ability resulting from Flashmeter training and gains in reading ability resulting from subsequent training in reading; (3) the transfer of training from one type of perceptual material to other types; and (4) changes in eye movements which occur with perceptual training.

* A condensation of a dissertation submitted in partial fulfillment of the requirements for the Ph. D. degree, in the Department of Psychology of the State University of Iowa. Dr. J. B. Stroud directed the study.

† Now at Bowling Green State University, Bowling Green, Ohio.

The experimental literature relating to the foregoing questions reveals a need for further experimentation before definite conclusions can be drawn. The relationship between perceptual span and reading ability is positive, as shown by Dallenbach,[8] Litterer,[28] Weber,[63] Gates,[16] Robinson,[35] Renshaw,[32] and Sutherland.[50] Correlation coefficients obtained in these studies range from .151 to .685. Dallenbach's correlations between perceptual span and school standing were .85 and .20, respectively, for a class of good students and a class of poorer students. The relationship between perceptual speed and reading ability is also positive, as shown by Weber,[63] Sister Mary,[33] Stroud,[48] and Renshaw.[32] Correlation coefficients obtained in these studies range from .016 to .867. No evidence was found as to the relationship between perceptual speed and perceptual span, although, as in Renshaw's study, one gets the impression that the relationship is positive. Both perceptual speed and perceptual span are positively related to reading ability, and there is room for doubt that perceptual speed and perceptual span are separate abilities. The experimental conditions under which they have been measured may be little more than two ways of measuring the same ability, whatever this ability be called.

Practice in the peripheral mechanics of reading, i.e., eye-movement training, can improve perceptual span. Data in support of this conclusion are provided by Pressey and Truax,[31] Robinson,[34] Witzeman,[66] Garver and Matthews,[15] Taylor,[52] Imus, Rothney and Bear,[24] Buswell,[6] and others. However, Sisson,[42] Buswell,[6] and Tinker[57] warn that eye-movement training does not necessarily bring about increased reading efficiency. Doubt is cast upon the causal nature of the relationship between eye-movement training and reading improvement as support for a peripheralistic view of the reading process when it is noted that comprehension scores do not differ significantly when learning gained through reading is compared with that gained from auditory presentation, when the results of error and vocabulary analysis in oral and silent reading are considered, and when comparisons are made of eye movements in oral and silent reading, as in the studies of Fairbanks,[12] Anderson and Swanson,[2] Anderson and Fairbanks,[3] Rogers,[36] Swanson,[51] and Buswell.[5] In addition, Tinker has found an insignificant relationship

between measures of ocular control and measures of reading ability.

The effect upon perceptual span of short-exposure types of training has been examined. Whipple[65] decided that practice did not improve perceptual span. However, Foster,[14] Dallenbach,[8,9] Gundlach, Rothschild, and Young,[22] Weber,[63,64] Seward,[39] Renshaw,[32] and Sutherland[50] report improvement in perceptual span with practice. Knehr[25] concluded that attempts to improve perceptual span invite failure in so far as peripheral factors are involved. Renshaw's results have been generally called into question by the results obtained by Army Air Corps psychologists using Renshaw's materials. Sutherland's measures of improvement in perceptual span are open to criticism on the basis of her scoring procedure. It would seem that the effectiveness of short-exposure training as a group method in improving perceptual span has not been established.

With respect to the effect of training in perceptual speed and in perceptual span upon subsequent performance in remedial reading courses, experimental evidence is slight. Most of the experimental work with perceptual speed and perceptual span has been directed toward discovering the relationship between them and reading ability. Renshaw[32] found that flash training with various materials increased reading rate and comprehension in his study. However, he did not conduct remedial reading training, and his enthusiastic claims regarding flash methods have yet to be substantiated. Sutherland[50] found that, with previous tachistoscopic training in perceptual span, her experimental group began remedial reading training at a higher level than did her control group, and made greater initial gains in reading ability. This superiority in amount of gain for the experimental group disappeared, however, midway through her remedial reading course, and in terms of gross gain the control group was significantly superior at the conclusion of the reading training. Sutherland's results are in need of testing with more reliable scoring methods. The establishment of a subject's perceptual span amounts to obtaining an upper limen for span, a procedure for which appropriate formulas are available. Sutherland did not use these formulas.

With respect to the prediction of improvability in reading from perceptual data, evidence is also slight. Tinker[57] has offered

evidence that measures of oculomotor efficiency, in terms of accuracy of voluntary fixation during successive perceptual acts and in terms of speed of convergence and divergence during successive visual fixations, do not correlate significantly with measures of reading speed or with photographic measures of reading ability. If variation in motor control of the eyes in voluntary fixation is not an important determinant of reading ability, prediction from this source would seem to be impossible. Litterer[28] concluded that visual apprehension scores, as indicated by the amount of material correctly reproduced after a three-second exposure, do not correspond highly with scores on reading tests. Stroud[48] found no predictive value for visual perception scores with respect to improvability in reading. However, he says, the abilities are related, and the possibility of prediction is not yet disproved. Knehr[25] concluded against any predictive value of perceptual span scores for reading performance. Robinson[35] says that perceptual span, as measured by the tachistoscope, is related to reading span, but probably should not be used to predict reading span.

The question of transference of the positive effects of perceptual training with one form of material to the ability to perceive other types of material seems still to be an open question. It raises the question of the generality of a perceptual factor. Whipple,[65] Foster,[14] Dallenbach,[8] and Litterer[28] conclude against the generality of a perceptual factor, although Litterer does admit the possibility. Gates[16] offers strong evidence against such generality. Sister Mary,[33] on the other hand, found the perceptual abilities examined in her study to be closely enough related to suggest a general perceptual factor, in Spearman's sense. Stroud[48] reports a significant relationship between rate of reading and rate of visual perception, involving digits, letters, words, and drawings. Renshaw[32] reports that tachistoscopic training with digits, English and Magyar words, and geometric forms all transferred to improve reading performance. However, it is not apparent from Renshaw's article exactly how these materials were presented; that is, whether the same groups received training with all types of material, or whether separate groups received training each with one type of material. If the former, it would be difficult to say to what extent transference from digits, say, improved reading rate, owing to the use also of English words as training material.

No evidence has been found bearing upon the question of the relationship between the degree of improvement in perceptual ability and the degree of improvement in reading ability. Sutherland's study could have provided data on this question, but she did not raise the question.

Similarly, no evidence was found regarding changes in eye movements accompanying training in perceptual speed and perceptual span. Robinson[35] correlated average fixation width as revealed by eye-movement photography during reading with tachistoscopic measures of perceptual span, finding a significant relationship. In this, Dearborn and Anderson[11] show agreement. However, Robinson gave no 'training' in perceptual span from which he could correlate changes in perceptual span with changes in eye-movement behavior. Other studies which have utilized eye-movement photography have been concerned with reading training, rather than with perceptual training.

This brief summary of the literature pertaining to the present study reveals a need for a well-controlled study which raises the questions of: (1) the relationships among perceptual speed, perceptual span, and reading ability; (2) the effect of perceptual training upon perceptual speed and perceptual span; (3) the effect of perceptual training upon reading ability; (4) the prediction of improvability in reading from perceptual data; (5) the generality of a perceptual factor, or transfer of practice effects from one perceptual material to another; and (6) the effect of perceptual training upon eye movements as measured by photographs taken during reading activity. The present study is an attempt to meet this need.

METHODS, PROCEDURE, APPARATUS AND MATERIALS

The subjects used in this study were freshmen in the remedial reading classes at the University of Iowa. With some exceptions, students in these classes have achieved scores corresponding to the 25th percentile or lower on the Blommers Test.[4] Ten such classes were available at the time the study was begun. Of these, six were chosen with the aid of a table of random numbers[27] (pp. 262–264). Two of the classes took perceptual speed training, two took perceptual span training, and two served as control groups. All the students in these classes ranked at or below the 25th percentile on the Blommers Test, except for seven who were at the 26th percentile, seven who were at the

29th percentile, and one who was at the 46th percentile. Complete records were obtained from one hundred ten subjects.

Two teachers took part in the study. Three of the six classes were assigned to each teacher by means of the table of random numbers, and each teacher's classes were assigned to method by the same means. Each teacher taught one perceptual speed training class, one perceptual span training class, and one control class.

Two methods of training were used in the attempt to improve perceptual ability. Both methods involved the flash presentation of words and phrases in a classroom situation. In the method called 'perceptual span training,' the words and phrases of the training series increased in length from six letters to thirty-four letters, in steps of two letters. In determining length of span, both letters and spaces between words were counted. Fifteen lengths of span were thus obtained. At each span there were forty-eight items, making a total of seven hundred twenty items in the training series. The exposure time for these items was 1/10th second. The schedule for their presentation comprises Table I.

TABLE I.—SCHEDULE FOR TRAINING THE PERCEPTUAL SPAN GROUPS

Day	1	2	3	4	5	6	7	8	9	10	11	12
	Span in letters and spaces											
	6*	8	10	12	16	18	20	22	26	28	30	32
	6	8	10	14	16	18	20	24	26	28	30	34
	6	8	12	14	16	18	22	24	26	28	32	34
	6	10	12	14	16	20	22	24	26	30	32	34
	8	10	12	14	18	20	22	24	28	30	32	34

* Each figure, in addition to denoting the span of the item, denotes one slide, containing twelve items. Thus, on Day 1, there were forty-eight items of a length of six letters and twelve items of a length of eight letters.

In setting up Table I, no consideration was given to the problem of establishing the optimal schedule. There is here a simple day-to-day progression in length of span presented. It is quite possible that some other arrangement would have been equally effective or more so.

In the method called 'perceptual speed training,' twelve-letter

words and phrases were presented according to a schedule of decreasing exposure times, from 1 to 1/100th second. There were seven hundred twenty of these items also. The short length was chosen in an attempt to avoid giving span training to the speed-training groups. This schedule appears in Table II.

TABLE II.—SCHEDULE OF SHUTTER SPEEDS FOR TRAINING THE PERCEPTUAL SPEED GROUPS

Shutter Speed (in seconds)	1	½	1/5	1/10	1/25	1/50	1/100
Days	Number of Items Presented*						
1	12	24	24				
2		24	24	12			
3		12	24	24			
4		6	18	24	12		
5			12	36	12		
6			12	24	12	12	
7				24	12	12	12
8				12	24	12	12
9				12	12	24	12
10					12	24	24
11					6	12	42
12					3	3	54

* Each number denotes a certain number of twelve-letter items. For example, on Day 1, there were twelve items presented at an exposure time of one second, twenty-four items at an exposure time of ½ second, and so on.

Table II was set up with the thought that very few items would be needed at the one-second exposure; that items at faster speeds should be introduced gradually; that proportionately more presentations should be made at middle speeds where presumably the integration of response patterns is difficult;[32] and that enough items should be given at the faster speeds to make practice at those speeds effective. It was not intended here to establish the optimal schedule.

As can be seen in Tables I and II, sixty items were presented at each class period for twelve class periods. Classes met four times a week. The subjects wrote each item on an answer sheet immediately after the item had been flashed upon the screen.

These answer sheets were graded in class for motivational purposes, but the scores were not used in the study.

In presenting the Flashmeter items, the instructors gave a ready signal, allowed one second to elapse, and exposed the item. The time between items was not constant, because of the time required for writing down the items.

Reading ability and perceptual ability were measured in all six classes at the start of the study by means of the following tests: the Blommers Rate of Comprehension Test; the Iowa Silent Reading Test (Advanced); the Thurstone Identical Forms

TABLE III.—SCHEDULE OF ACTIVITIES FOR EXPERIMENTAL AND CONTROL GROUPS

Class Periods	Perceptual Speed Groups	Perceptual Span Groups	Control Groups
	Activity		
1–4	Pretests of perceptual and reading ability.*	Pretests of perceptual and reading ability.*	Pretests of perceptual and reading ability.*
5–16	Training in perceptual speed.	Training in perceptual span.	Class work in speaking and writing.
17–20	Post-tests of perceptual and reading ability.	Post-tests of perceptual and reading ability.	Post-tests of perceptual and reading ability.
21–32	Intensive reading training.	Intensive reading training.	Intensive reading training.
33–34	Final test of reading ability.	Final test of reading ability.	Final test of reading ability.

* See *Criteria*. The same tests, or alternate forms, were used as pretests and as post-tests. The Iowa Silent Reading Test was not given during the testing on days 17–20; it was used on days 33–34. The Blommers Test was given during the testing on days 17–20, but not on days 33–34. Both of these tests were given as pretests.

Test; four subtests of the I.E.R. Test, C2, Form B; four subtests of the I.E.R. Test, C1; and Flashmeter tests of perceptual span (words and phrases), perceptual speed (words and phrases)

and digit span. Eye-movement records were made with the Ophthalm-O-Graph.

After the administration of the initial battery of tests, the training for the experimental groups took place, as described above. During the same time, the two control groups carried on work in speaking and writing in fulfillment of other requirements of their regular courses. Following this period, which occupied three weeks, the second battery of tests was administered to all six classes. This battery comprised the same tests as the original battery, or alternate forms, except that the Iowa Silent Reading Test was not administered at this time. All six classes then entered upon a period of intensive reading training, using the text of the Harvard Reading Films printed in pamphlet form, and equated transfer reading selections currently in use in University of Iowa remedial reading classes. Ten-item multiple-choice comprehension tests were given after the reading of each selection in either series of reading material. Some vocabulary work and exercises in critical reading were also undertaken during this time, with the assignments being the same in all six classes. This intensive reading training period extended over three weeks, with four class meetings each week.

At the conclusion of the reading training period, the Iowa Silent Reading Test (Advanced) was again administered (alternate form) to all six classes.

The schedule of activities just described is summarized in Table III.

Apparatus.—The American Optical Company's Ophthalm-O-Graph and the Keystone Overhead Projector with Flashmeter attachment were used in this study. The Ophthalm-O-Graph is an efficient eye-movement camera, by means of which the movements of points of light reflected from the corneas of the eyes are recorded upon motion picture film during reading, and scores for reading rate, number and duration of fixations, number of regressions, and reading span can be derived.

The Overhead Projector is so constructed as to permit any desired manipulation of the slide. By means of cut-out masks, any portion of the slide may be projected. In the present study, a mask cut to permit the projection of a single line was used. It was thus possible to have twelve items on each slide, although only one at a time was projected. The Flashmeter attachment

is a large camera shutter having speeds from 1 to 1/100th second, which fits over the lens barrel of the projector. Throughout the study, the projector was placed fourteen and one-half feet from the screen, measured to the base of the projector. A yellow filter over the object lens cut down the brightness differential between the items and their backgrounds. The image thrown onto the screen was quite brilliant and clear, and it was possible to allow some extra light from the windows so that the subjects could see to write the items after each presentation. The screen image was clearly visible from any part of the room.

Materials.—The Flashmeter tests of perceptual span and perceptual speed and the training materials for the experimental groups were presented on standard projection slides by means of the Flashmeter. The words and phrases used in making the slides were cut from a freshman reading textbook, pasted on white cards in groups of twelve, and slides were made from the cards by the usual photographic method.

Criteria for the selection of the words and phrases were the same as those used by Sutherland.[50]

The reading material used for intensive reading practice was a printed form of the Harvard Film text. There are fifteen films, but time permitted using only the first twelve. In addition to these printed selections, the Harvard equated transfer reading selections were used. The printed form of the film text and the equated transfer selections alike are accompanied by a ten-item multiple-choice comprehension test for each selection. In each daily reading practice period, one printed Harvard Film text booklet and one Harvard equated reading selection were read, and the comprehension test taken over each. The subjects kept a record of their progress during the intensive reading practice periods.

Criteria.—Twenty-seven criterion measures were obtained in this experiment. Five were obtained from Ophthalm-O-Graph records, while the other twenty-two came from the administration of standardized reading tests, Flashmeter tests, parts of two standardized clerical tests, and the Thurstone Identical Forms Test. The Flashmeter tests were included to discover to what extent perceptual span, perceptual speed, and digit span, as measured by the Flashmeter, will improve with Flashmeter training. The reading tests were included to discover

what effect perceptual training given with the Flashmeter might have upon reading performance. The Ophthalm-O-Graph records were analyzed to see if changes in perceptual ability brought about by Flashmeter training were accompanied by changes in eye movements, as revealed by eye-movement photographs, and whether the two flash methods differed in this respect. The clerical tests (which included number cancellation, use of tables, and reading backward) and the Thurstone test were included to see whether or not Flashmeter training in perceptual span and perceptual speed transferred to other perceptual materials not involving flash presentation.

Flashmeter Tests:—Three Flashmeter tests were used, one for perceptual span for words and phrases, one for digit span, and one for perceptual speed involving words and phrases. There were two forms of each test, devised by the writer in the manner described earlier in the present paper.

In each form of the test of perceptual span for words and phrases, there were ninety items, presented six each at fifteen spans of from six to thirty-four letters, counting letters and spaces between words. Each span was two letters longer than the next smaller span. All items were presented at a shutter speed of 1/10th second. As in the training series, the projector was placed fourteen and one-half feet from the screen. The test conditions approximated those of the training periods, although the subjects did not grade their own papers as they did in the training periods.

The test of perceptual span for digits was also presented under these conditions. Spans of from two to sixteen digits were presented, with each span being larger by two digits than the next smaller span. There were forty-eight items in each form of the digit span test.

In the test for perceptual speed for words and phrases, the items were twelve letters in length. They were presented at a shutter speed of 1/100th second. Each form of the test comprised forty-two items.

The perceptual speed test was scored in terms of the number of items correctly reproduced after exposure. The perceptual span tests were scored according to Spearman's arithmetical method for computing limens[67] (pp. 400–404). In all the Flashmeter tests, the word, phrase, or series of digits had to be written

exactly as shown in order to be counted as correct. No partial credits were given for items correctly reproduced in part.

A reliability coefficient of .86 was obtained for the Flashmeter test of perceptual span for words and phrases, with subjects not involved in the study. There is no estimate of the reliability of the other Flashmeter tests used.

The Iowa Silent Reading Test, Advanced[21]:—This test consists of seven parts, two of which have two subparts, and so nine scores are obtained. Form Cm was used as the initial test, and Form Dm as the final test. The regular time limits were used.

The Blommers Test[4]:—In this test there are twenty independent reading exercises, with a question to be answered over each. The score is based upon the mean length of time taken to read and answer only those exercises correctly answered. Form B was used as an initial test, and Form A as a final test. The regular time limits were used.

The Thurstone Identical Forms Test[53]:—This is a test of perceptual speed, in which the subject selects from five formalized designs the one exactly like a sample figure. There are sixty items of this kind. The score is the number of items correctly marked. The usual time limit is eight minutes, but only four minutes were allowed in the present study, to insure the possibility of better scores being made on the retest. Since there is only one form of this test, it had to be used at both initial and final testing. Practice effects could be estimated from the performance of the control classes.

I. E. R. Clerical Tests, Form C1 and C2, Form B:—These tests, issued by the Institute of Educational Research, Teachers College, Columbia University, are tests of clerical abilities, of which some are undoubtedly of a perceptual nature. In Form C1, subtests 4, 7, 9, and 10 were used. In Form C2, subtests 1, 2, 3, and 4 were used. Other parts of the tests were not used because they ostensibly involve skills other than perceptual ability, such as mathematical ability, psychomotor skill, or general knowledge. Since only one form of each of these tests was available, they were given both as pretests and as post-tests. The possible practice effect of this procedure could be estimated from the performance of the control groups. Time limits for the parts used were: C1, part 4, four minutes; C1, part 7, three

minutes; C1, part 9, three minutes; C1, part 10, three minutes; C2, part 1, two minutes; C2, part 2, one and one-half minutes; C2, part 3, two minutes; C2, part 4, three minutes. These times, shorter than standard for these tests, were used to insure the possibility of higher scores being made upon the second administration of the tests.

Eye-movement Records:—The test selection used in obtaining these records was that used by Glock[17] and Stone.[44] There were three hundred words, set in 12-point type. Average length of word was three ems. Since it has been shown[40,44] that the most efficient measures of eye movements in a 300-word passage are to be had from the last hundred words, eye movements were photographed only during the reading of the third hundred words. Average length of line in the last hundred words of the test selection was twenty-nine ems. Each subject read one passage before the camera before he read the test selection. The sound of the camera motor was masked by a second motor, so that the subjects did not know when the records were being made.[55] Comprehension tests (from Stone[44]) were given following the reading of the test selection. The same selection was used in initial and final testing, in view of the finding of Imus and Rothney[24] that no greater variability in eye movements is shown under these circumstances than is shown when different selections are read at initial and final testing.

The records obtained from the Ophthalm-O-Graph included the number of fixations and regressions (in counting fixations, regressions were counted as fixations), the average span of recognition in ems, the average duration of fixations in seconds, and reading rate in ems per minute. Duration of fixation was found by dividing the total time of reading by the total number of fixations. Although total reading time includes the time taken for return sweeps to the beginnings of lines, this time for return movement was disregarded, in view of Tinker's finding[56] that return movements take less than ten per cent of total reading time. Tinker[55] has reported reliabilities ranging from .72 to .88 for similar measures.

RESULTS

The Relationships Among Perceptual Speed, Perceptual Span, and Reading Ability.—The test battery yielded perceptual span

and perceptual speed scores (Flashmeter tests) and scores on reading 'rate of comprehension' (Blommers Test) and on nine subtests of the Iowa Silent Reading Test. Product-moment correlation coefficients were computed between initial scores on the two Flashmeter tests, and between initial scores on these tests and initial scores on the reading tests, for all subjects (N = 110). The results are given in Table IV.

TABLE IV.—THE RELATIONSHIPS AMONG INITIAL TEST SCORES IN PERCEPTUAL SPEED, PERCEPTUAL SPAN, AND READING

	Perceptual Span	P*	Perceptual Speed	P*
Blommers Test	.200	5%	.183	N.S.
Iowa Silent Reading Test				
Rate	.179	N.S.	.162	N.S.
Comprehension	.123	N.S.	.128	N.S.
Directed Reading	.124	N.S.	.046	N.S.
Poetry Comprehension	.183	N.S.	.130	N.S.
Word Meaning	.209	5%	.212	5%
Sentence Meaning	.164	N.S.	.191	5%
Paragraph Comprehension	.292	1%	.251	1%
Use of the Index	.316	1%	.146	N.S.
Selection of Key Words	.217	5%	.183	N.S.
Perceptual Speed	.758	1%		

* Computed according to Lindquist[27] (pp. 210 ff.).

Of special interest is the high relationship obtained between perceptual span and perceptual speed. Not only is this coefficient statistically significant, but it is high enough to suggest that perceptual span and perceptual speed, as measured with the tachistoscope, are not very greatly differing abilities.

The present study agrees with previously published studies in finding a positive relationship between perceptual span and reading ability, although only five of the obtained coefficients are significant. The obtained relationships between perceptual speed and reading ability are also positive, although only three of the obtained coefficients are statistically significant. One might expect, on *a priori* grounds, that the relationships mentioned would be positive, and other studies have agreed in finding

them so. However, the coefficients obtained in several other studies are much larger than those obtained in the present study, ranging as high as .685 between perceptual span and reading ability, and as high as .867 between perceptual speed and reading ability. The relationships found in the present study may well be a function of the restricted range of reading ability of the subjects used, rather than an approximation of 'true' relationships among these variables.

The Effect of Flashmeter Training upon Performance on Flashmeter Tests.—Before an evaluation was made of the effects of Flashmeter training upon performance in other types of tests, it seemed desirable to see whether or not the Flashmeter training brought about any significant changes in performance on the Flashmeter tests themselves. In this connection, the differences between initial and final scores on the Flashmeter tests were evaluated by means of the *t*-test. The results of this analysis are presented in Table V.

It can be seen that except in one case the gains made from initial to final Flashmeter testing were significant at the one-per-cent level for both perceptual span and perceptual speed training, on both tests, and for both teachers. Not only did all the experimental groups make significant gains in the functions practiced, but they also made significant gains in the functions not practiced, with one exception. This finding concurs with the finding (Table IV) of a high relationship (.758) between initial scores on the Flashmeter tests of perceptual span and perceptual speed.

Because Teacher A's control group made a gain on the perceptual speed test which was significant at the one-per-cent level, it was decided to analyze these data by means of factorial analysis of variance, correcting for differences in class size in accordance with Lindquist[27] (p. 130). Table VI presents the results of this analysis.

It can be seen that the perceptual span test shows an interaction variance which is significant at the one-per-cent level, whereas the methods variance is not significant. It would appear that the relative effectiveness of the methods is not the same for the two teachers involved, and that the experimental groups did not improve in perceptual span significantly more than did the control groups.

TABLE V.—SIGNIFICANCE OF THE DIFFERENCES BETWEEN INITIAL AND FINAL SCORES ON FLASHMETER TESTS OF PERCEPTUAL SPAN AND PERCEPTUAL SPEED

	Initial Mean	Final Mean	Mean Gain	$6\sigma_M$	t	P
Teacher A, Speed Training Group (N 18)						
Perceptual Span Test	15.089	15.962	.873	.526	1.661	20%
Perceptual Speed Test	26.000	32.667	6.667	1.889	3.529	1%
Teacher A, Span Training Group (N 13)						
Perceptual Span Test	14.202	16.718	2.516	.820	3.069	1%
Perceptual Speed Test	25.462	32.539	7.077	2.249	3.147	1%
Teacher A, Control Group (N 17)						
Perceptual Span Test	14.095	15.216	1.121	.599	1.872	10%
Perceptual Speed Test	24.412	28.235	3.823	1.148	3.330	1%
Teacher B, Speed Training Group (N 21)						
Perceptual Span Test	13.396	16.365	2.969	.590	5.028	1%
Perceptual Speed Test	19.286	29.476	10.190	1.710	5.959	1%
Teacher B, Span Training Group (N 20)						
Perceptual Span Test	13.565	16.068	2.503	.752	3.326	1%
Perceptual Speed Test	22.500	30.400	7.900	1.765	4.463	1%
Teacher B, Control Group (N 21)						
Perceptual Span Test	13.803	13.144	−.659	.494	1.334	20%
Perceptual Speed Test	20.857	24.619	3.762	1.814	2.074	10%

Standard error of the Mean Gain ($6\sigma_M$), *t*-ratio (t), level of significance (P).

TABLE VI.—FACTORIAL ANALYSIS OF VARIANCE, FLASHMETER TESTS OF PERCEPTUAL SPAN AND PERCEPTUAL SPEED

Test	F Interaction	P	F Methods	P	F Teacher	P
Perceptual Span	4.83	1%	1.73	N.S.*	.01	N.S.*
Perceptual Speed	.52	N.S.	4.21	5%	1.24	N.S.

* Interaction variance used as error term.

For the perceptual speed test, the methods variance is significant at the five-per-cent level, whereas the interaction variance is not significant. The implication is that there is a reliable methods difference appearing on this test. Application

of the *t*-test to the mean gains on this test reveals that for Teacher A, neither the difference between the speed training group and the control group, nor that between the span training group and the control group is large enough to be significant. For Teacher B, however, the difference in mean gain between the speed training group and the control group is significant at the one-per-cent level, and that between the span training group and the control group is significant at the five-per-cent level. Since the interaction variance was not significant, the relative effectiveness of the methods may be assumed not to differ significantly from one teacher to the other, and the insignificant teacher variance indicates that the difference between teachers is not significant. It may be said, then, that either method of Flashmeter training, perceptual span or perceptual speed, produced a reliably greater mean gain in perceptual speed than was made by the control groups, at the five-per-cent level of confidence.

Sutherland[50] found a gain in perceptual span with tachistoscopic training which was significant at the one-per-cent level. It should be noted, however, that her control groups did not take the span tests, and so her evaluation of the gains made by her experimental groups is incomplete. Renshaw[32] also used no control groups in evaluating the improvement in reading shown by his Battelle Memorial Institute and Arlington subjects after tachistoscopic training with several kinds of materials.

That these studies secured improvement in performance on tests of the same ability as that involved in training is not being questioned. Rather, it seems desirable to find some reason for the finding in the present study that, although the experimental groups made significant gains in the measures used, the control groups gained enough to render the methods variance insignificant for the perceptual span test, and to make the methods variance just significant at the five-per-cent level for the perceptual speed test. Either the Flashmeter training was ineffective (but Table V shows that it was effective) or the control groups in some way obtained enough practice to bring about the results noted. That the latter possibility is likely under the conditions of the present study is suggested by an earlier study conducted by Dallenbach.[8]

Dallenbach's conditions were not strictly comparable to those

of the present study, but they were similar. He conducted short-exposure training upon school children, using a variety of materials. He evaluated the progress of his subjects by means of five tests, of which the first four were two weeks apart. During each of these four-week periods, he presented two hundred items to his subjects. The curve of improvement which he obtained showed a marked negative acceleration, with the most marked improvement being shown at the second test, or after the administration of two hundred items, and with the curve being quite flat thereafter. Foster[14] and Seward[39] have made similar observations.

To explain the results reported just above, from the present study, it is suggested that the initial Flashmeter tests were long enough to give the control groups a considerable degree of practice toward bettering their performance on the later Flashmeter tests; i.e., enough practice to result in a gain comparable to that represented by the steep initial portion of Dallenbach's curve. The initial Flashmeter tests used in the present study comprised a total of one hundred eighty items (ninety items in the span test, forty-two items in the speed test, and forty-eight items in the digit span test), or very nearly the same number of items as were presented to Dallenbach's subjects during their initial training period. In addition, the possibility must not be overlooked that practice on early items of the final Flashmeter tests had some influence upon total performance on the final Flashmeter tests. (The tests were long because reliable tests were desired). It is possible, then, that the training given to the experimental groups in the present study affected only that part of improvement in performance which is represented by the relatively flat portion of a negatively accelerated curve of improvement. It is recognized, *a posteriori*, that giving the initial Flashmeter tests to the control groups constitutes a complication in the interpretation of the obtained results. However, in view of the discussion just concluded, the results obtained in the present study may be brought into general agreement with those of previously published studies.

The Effect of Flashmeter Training in Perceptual Speed and Perceptual Span Upon Scores in Reading Tests.—The second major question raised in the present study concerns the effect of Flashmeter training in perceptual span and perceptual speed

upon performance on reading tests. Under the design of the study, this question may be thought of as having two parts: (1) What is the effect of the Flashmeter training itself upon reading performance, and (2) What effect does the Flashmeter training have upon the degree of improvement shown after a short and intensive reading practice period? Before Flashmeter training began, reading ability was measured by the Blommers Test (Form B) and the Iowa Silent Reading Test (Form Cm). After Flashmeter training had been given, reading ability was again measured by means of the Blommers Test (Form A). The data obtained from the two administrations of the Blommers Test were applied to the first part of the question now under consideration. At the end of the three-week intensive reading-practice period, reading ability was measured a third time, by means of the Iowa Silent Reading Test (Form Dm). The data obtained from the two administrations of this test were applied to the second part of the present question. The gains made on both tests were analyzed by means of factorial analysis of variance, and the significance of the gains was determined by means of the *t*-test. In the factorial analysis of variance, correction was made for differences in class size according to Lindquist,[27] (p. 130). Table VII contains the results of the factorial analysis of variance.*

The mean gains and *t*-ratios for the Blommers Test have been corrected for regression toward the population mean. Since the subjects were enrolled in these classes because of low initial scores on the Blommers Test, there should be some increase in scores with a readministration of the test even though no instruction of any sort intervened. However, the corrected *t*-ratios between initial and final Blommers Test scores were significant for all but one class.

A look at Table VII reveals that the methods variance for the Blommers Test is not significant. Apparently there is no

---

* Because of space limitations, several tables have been omitted. These include: "Significance of the Mean Gains from Initial to Final Testing on the Blommers Test and the Iowa Silent Reading Test," "Significance of the Mean Gains from Initial to Final Testing on Perceptual Tests Other than the Flashmeter Tests," and "Significance of the Differences between Initial and Final Ophthalm-O-Graph Measures." Copies of these tables may be obtained by writing to the author.

TABLE VII.—FACTORIAL ANALYSIS OF VARIANCE ON GAINS MADE FROM INITIAL TO FINAL TESTING WITH THE BLOMMERS TEST AND THE IOWA SILENT READING TEST

	F Interaction	P	F Methods	P	F Teacher	P
Blommers Test	1.55	N.S.	.52	N.S.	8.70	1%
Iowa Silent Reading Test						
Rate	40.92	1%	.16	N.S.*	.59	N.S.*
Comprehension	2.23	N.S.	.30	N.S.	.32	N.S.
Directed Reading	.04	N.S.	1.79	N.S.	5.74	5%
Poetry Comprehension	4.14	5%	1.61	N.S.*	.11	N.S.*
Word Meaning	1.18	N.S.	.03	N.S.	.89	N.S.
Sentence Meaning	4.83	1%	.54	N.S.*	1.21	N.S.*
Paragraph Comprehension	2.16	N.S.	1.76	N.S.	12.03	1%
Use of the Index	.90	N.S.	3.26	5%	.64	N.S.
Selection of Key Words	2.18	N.S.	.29	N.S.	.67	N.S.

* Interaction variance used as error term.

difference in the relative effectiveness of the methods of training, with respect to performance on this test. In view of the discussion in the previous section, this result might have been expected. If the hypothesis offered there is correct, the training received by the control groups from the administration of the Flashmeter tests was sufficient to account for the significance (one-per-cent level) of the corrected gains made by the control groups on the Blommers Test.

With respect to the subtests of the Iowa Silent Reading Test, the gains made by all the groups were significant in the great majority. This was to be expected, if the reading practice intervening between the Flashmeter training and the final testing with the Iowa Silent Reading Test was itself effective. Glock[17] found that reading practice with the printed form of the Harvard Films did improve reading performance. The results substantiate his finding. However, the present study is concerned also with the question of whether or not the Flashmeter training brought about significantly greater improvement in reading performance on the part of the experimental groups than was made by the control groups. The factorial analysis of variance made on the gains shown on the Iowa Silent Reading Test (Table VII) reveals that only one subtest yielded a sig-

nificant methods variance. The subtest is Use of the Index, and the level of significance is five per cent. The mean gain made by Teacher A's span-training group exceeds that made by his speed-training group by a significant amount, as shown by the *t*-test, and the same is true of his control group. The difference in mean gain between Teacher A's span-training and control groups is not significant. For Teacher B, none of the differences among groups is large enough to be significant. For both teachers, the control group made a greater mean gain than was made by the speed-training group. Apparently, then, it cannot be said that the experimental groups made significantly greater gains than were made by the control groups.

The gains made on the Blommers Test were generally significant, and the methods differences were not significant (Table VII). The same is true of the Iowa Silent Reading Test. Since the factor which differed from the experimental to the control groups was the Flashmeter training given to the experimental groups, it continues to appear likely that the practice obtained by the control groups from the Flashmeter tests played a large part in the appearance of the results thus far presented.

Prediction of Improvability in Reading from Perceptual Data.— The question here is really that of whether or not there is any relationship between initial Flashmeter scores and the amount of reading improvement accompanying Flashmeter training. It was found that significant gains in reading performance do follow Flashmeter training, as shown by the Blommers Test. To see if these gains in reading performance could have been predicted from initial Flashmeter test scores, correlation coefficients were computed between initial Flashmeter test scores and the gains made on the Blommers Test. This was done according to Lindquist,[27] (pp. 219 ff.), using the method for computing the correlation within schools; here, classes were substituted for schools, and the obtained coefficients may be interpreted as averages of the coefficients which would be found for the separate classes. The coefficient obtained between the Blommers gain score and the initial Flashmeter perceptual speed test score was .014; that between the Blommers gain score and the initial Flashmeter perceptual span test score was −.153. These coefficients are not significant, and are so low that they have no predictive value.

Relationship Between Improvement in Performance on Flashmeter Tests as a Result of Flashmeter Training and Improvement in Reading Resulting from Reading Training.—Do subjects who improve greatly in perceptual ability by virtue of having undergone Flashmeter training improve in reading through reading training more than do subjects who improve little in perceptual ability with Flashmeter training? In pursuit of this question, correlation coefficients were computed between the gains made from initial to final Flashmeter testing and the gains made from initial to final testing with the Iowa Silent Reading Test. The computations were made according to Lindquist[27] (pp. 219 ff.). The following coefficients were obtained.

	Flashmeter Speed Test Gain Score	Flashmeter Span Test Gain Score
Gain Score on the Iowa Silent Reading Test,		
Rate	.176	−.166
Comprehension	.141	−.207*
Directed Reading	−.174	−.051
Poetry Comprehension	.004	−.222*
Word Meaning	−.049	−.065
Sentence Meaning	.035	−.076
Paragraph Comprehension	.024	−.066
Use of the Index	−.089	−.071
Selection of Key Words	.088	−.096

* Significant at the five-per-cent level.

None of the coefficients of correlation between the gains on the Flashmeter test of perceptual speed and the Iowa Silent Reading Test gain scores are significant. Two of those between the Iowa Test and the Flashmeter span test gain scores are significant at the five-per-cent level. The relationship between the span test gain scores and gain scores on the subtests of the Iowa Reading Test seems to be negative, although the coefficients are all so low that it is doubtful that any significant trend is revealed.

Transfer of Flashmeter Training to other Perceptual Materials.—The question raised here is really one of the generality of a perceptual factor; as has been shown in the introduction of the

present paper, this question is far from being settled. Perceptual tests other than the Flashmeter tests were given both before and after the Flashmeter training in the present study, to see to what extent Flashmeter training transferred to performance on these other perceptual tests. The significance of the gains made on these tests was assessed by means of the *t*-test, and factorial analyses of variance were made on the gain scores to see whether the experimental groups, who received Flashmeter training, made significantly greater gains than were made by the control groups, who had no such training ('no such training' means that the control groups underwent no planned Flashmeter training; the possible effect upon the performance of the control groups of the initial Flashmeter tests must be kept in mind).

The results of the *t*-test were generally significant—it would appear that there is some positive transfer effect from the Flashmeter training in perceptual span and perceptual speed to other perceptual materials, and that some measure of support is given to the idea of the generality of a perceptual factor. The performance of the control groups, for whom the *t*-ratios are also predominantly significant, suggests two possibilities: (1) the earlier hypothesis as to the effect on the control group's performance of the initial Flashmeter tests is further supported, or (2) since the same form of the tests were used at initial and final testing (except the Digit Span test), it is possible that the significant results shown here are owing to a practice effect. Under the design of the present study, it could not be determined which possibility is the more likely. The results of the factorial analysis of variance are presented in Table VIII.

A *t*-test between gains made by the control and experimental groups on the three tests yielding significant methods variances shows the Flashmeter training in perceptual span to be superior to Flashmeter training in perceptual speed in transferring to improve performance with other perceptual materials. The differences are significant at the five-per-cent level of confidence.

The Effect of Flashmeter Training on Eye Movements Made while Reading.—Eye-movement records were made before and after Flashmeter training, using the Ophthalm-O-Graph. The significance of the changes from initial to final photographing was determined by means of the *t*-test, and methods differences were evaluated by means of factorial analysis of variance using

TABLE VIII.—FACTORIAL ANALYSIS OF VARIANCE ON PERCEPTUAL TESTS OTHER THAN THE FLASHMETER TESTS OF PERCEPTUAL SPAN AND PERCEPTUAL SPEED

Test		F Interaction	P	F Methods	P	F Teacher	P
Thurstone Identical Forms		10.11	1%	.27	N.S.	.50	N.S.*
Digit Span, Flashmeter		3.11	5%	.49	N.S.	7.48	1%*
IER, C2, Form B, Test 1		1.55	N.S.	4.26	5%	.03	N.S.
	Test 2	2.18	N.S.	1.78	N.S.	.40	N.S.
	Test 3	2.97	N.S.	.70	N.S.	.03	N.S.
	Test 4	.38	N.S.	4.13	5%	2.26	N.S.
IER, C1	Test 4	.49	N.S.	4.73	5%	2.15	N.S.
	Test 7	1.79	N.S.	1.36	N.S.	.66	N.S.
	Test 9	.97	N.S.	.45	N.S.	5.90	5%
	Test 10	2.66	N.S.	.03	N.S.	2.89	N.S.

* Interaction variance used as error term.

the gains as data. There is no very consistent trend toward significance in the changes in any of these measures, with the exception of reading rate in ems per minute. In view of results previously discussed, this might have been expected; i.e., the finding of generally significant changes on the Blommers Test and the Iowa Silent Reading Test. It should also be noted that changes in other Ophthalm-O-Graph measures, while generally not significant, are predominantly in the direction of gains. The cumulative effects of these small gains undoubtedly contribute to the significance of the changes in rate of reading.

The results of the factorial analyses of variance made on the gain scores in the Ophthalm-O-Graph measures are given in Table IX. This table shows that methods variances significant at the five-per-cent level were obtained for two of the eye-movement measures. One of these measures is the reduction in number of regressions per three hundred ems. Neither the interaction variance nor the teacher variance was significant for this measure. Comparison of the gains from group to group by means of the *t*-test revealed that Teacher A's experimental groups both gained significantly more than his control group (at the one-per-cent level), with the gains for the experimental groups not differing to a significant degree. Teacher B's span group gained significantly (five-per-cent level) over her control group. Her

TABLE IX. FACTORIAL ANALYSES OF VARIANCE ON GAINS IN EYE-MOVEMENT MEASURES

Ophthalm-O-Graph Score	F Interaction	P	F Methods	P	F Teacher	P
Reduction in no. of fixations/300 ems	1.68	N.S.	1.49	N.S.	.07	N.S.
Reduction in no. of regressions/300 ems	.48	N.S.	4.34	5%	.24	N.S.
Increase in Rate, in ems per minute	3.35	5%	.84	N.S.	.001	N.S.*
Reduction in duration of fixation, in seconds	5.60	1%	.007	N.S.	.40	N.S.*
Increase in recognition span, in ems	.02	N.S.	3.36	5%	.01	N.S.

* Interaction variance used as error term.

speed group did not gain significantly more than did her control group, nor were the gains for her speed and span groups significantly different. With respect to recognition span in ems, a *t*-test comparing gains among groups showed that both of Teacher A's experimental groups gained significantly more (five-per-cent level) than his control group, with the difference in gain between the experimental groups being not significant. Teacher B's span group gained significantly more (five-per-cent level) than did her control group, but her speed group did not. The difference in gain between Teacher B's experimental groups is not significant. Both span-training groups made the largest mean gain in recognition span.

Doubt is cast upon the effectiveness of Flashmeter training in perceptual span and perceptual speed in influencing eye-movement measures, as compared with other remedial reading methods, when the gains made by Glock's subjects[17] are considered. On the two eye-movement measures showing significant methods variances in the present study, the mean gains for Glock's classes ranged from 25.07 to 37.55 for reduction in number of fixations per three hundred ems, and from 1.22 to 1.94 for increase in recognition span in ems. On the Ophthalm-O-Graph measure of reading rate, the gains for Glock's classes range from 346.22 to 654.68 ems per minute. The present study obtained mean gains for the experimental groups of from 3.37 to 9.77 for reduction in number of fixations per three hundred ems,

and from .161 to .309 for increase in recognition span in ems. On the Ophthalm-O-Graph measure of reading rate, the gains for the groups used in the present study ranged from 49.18 to 135.13 ems per minute. Compared with Glock's gains, obtained with the use of remedial reading films and printed practice reading material over a similar period of time, the gains obtained in the present study are rather small.

CONCLUSIONS

It should be remembered that the subjects used in this study were of a restricted range of reading ability, as measured by the Blommers Test. The application of the results here presented to other groups not similarly restricted should be made with care, if at all. In addition, the results obtained may be purely a function of the use of the Flashmeter in measuring perceptual span and perceptual speed and in giving training in these functions. Generalization of the obtained results to other methods of measuring and training these functions should be done with caution. Subject to the conditions of the study, the following conclusions seem to be warranted.

1) Perceptual span and perceptual speed, as measured by the Flashmeter, are positively related; the coefficient is high (.758). Correlation coefficients obtained between Flashmeter measures of perceptual span and reading test scores are positive but low; the same is true of coefficients obtained between reading test scores and Flashmeter measures of perceptual speed.

2) Significant gains were made from initial to final Flashmeter testing, not only in the function practiced (perceptual span or perceptual speed) but also in the one not practiced. However, neither training method produced gains significantly greater than those made by the control groups. Evidence was offered to support the suggestion that the length of the Flashmeter tests was sufficient to produce marked gains on the part of the control groups.

3) As measured by the Blommers test, significant gains were made in reading performance, from the first administration of the Blommers test to the second. Flashmeter training was given to the experimental groups in this interval. However, the insignificant methods variance obtained on this test (as shown in Table VII) leads to the conclusion that the experimental groups

did not improve in reading performance more than did the control groups. This result adds weight to the suggestion that the length of the initial Flashmeter tests was sufficient to affect the reading performance of the control groups.

4) As measured by the Iowa Silent Reading test, significant gains were made in reading performance after the Flashmeter training period plus the period of intensive reading practice. Again, the methods differences were not significant. This result further supports the suggestion that the control groups received a substantial amount of perceptual training from the Flashmeter tests.

5) Prediction of reading improvement from initial Flashmeter test scores was not possible on the basis of the results obtained in the present study.

6) The relationship between gains in perceptual ability resulting from Flashmeter training and gains in reading ability resulting from Flashmeter training plus reading training is small and not significant.

7) Gains in perceptual tests other than the Flashmeter perceptual span and perceptual speed tests are generally significant. The three tests showing significant methods variances agree in finding training in perceptual span to be superior to training in perceptual speed in transferring to improve performance with other perceptual materials. If the suggestion as to the effect of the long initial Flashmeter tests is acceptable, these results may be interpreted as supporting the notion of a general perceptual factor.

8) Except for rate of reading, changes in eye-movement measures were generally insignificant, although there is general improvement in all eye-movement measures. The gains made in the present study, using Flashmeter training, are small in comparison to the gains made by other training methods used in other studies.

9) Of twenty-seven factorial analyses of variance made in this study, twelve show significant teacher variances or significant interaction variances, or both. Provision for evaluating the teacher variable as an aid in interpreting the results of methods experiments is an essential precaution, and one which has been overlooked in some studies, such as those of Renshaw[32] and Sutherland.[50]

10) The smallness of the gains made in the present study suggests that the methods used were not as effective as other training methods which have been used in other studies. The fault may lie with Flashmeter training itself, or it may be that group methods of training are generally less effective than individual methods. Certainly there is the possibility that training methods other than those used in this study might secure greater improvement in perceptual speed and span, and thus have greater effect upon reading performance.

BIBLIOGRAPHY

1) I. H. Anderson, "Studies in the Eye Movements of Good and Poor Readers," *Psychological Monographs*, 1937, 48, 1–35.

2) I. H. Anderson and D. E. Swanson, "Common Factors in Eye Movements in Silent and Oral Reading," *Psychological Monographs*, 1937, 48, 61–69.

3) I. H. Anderson and G. H. Fairbanks, "Common and Differential Factors in Reading Vocabulary and Hearing Vocabulary," *Journal of Educational Research*, 1937, 30, 317–332.

4) P. J. Blommers, "Rate of Comprehension of Reading: Its Measurement and Its Relation to Comprehension," Ph. D. Thesis, State University of Iowa, Iowa City, 1944.

5) G. T. Buswell, "Fundamental Reading Habits, A Study of Their Development," *Supplementary Educational Monographs*, Chicago: University of Chicago Press, No. 21, 1922.

6) ———, "Remedial Reading at the College and Adult Level," *Supplementary Educational Monographs*, Chicago: University of Chicago Press, No. 50, 1939.

7) Luella W. Cole, *The Improvement of Reading with Special Reference to Remedial Instruction*, New York: Farrar and Rinehart, 1938.

8) K. M. Dallenbach, "The Effect of Practice upon Visual Apprehension in School Children," *Journal of Educational Psychology*, 1914, 5, 321–344, 369–404.

9) ———, "The Effect of Practice upon Visual Apprehension in the Feeble-Minded," *Journal of Educational Psychology*, 1919, 10, 61–62.

10) W. F. Dearborn, *Psychology of Reading*, Columbia University Contributions in Philosophy and Psychology, 1906, 14, No. 1.

11) W. F. Dearborn and I. H. Anderson, "A New Method for Teaching Phrasing and for Increasing the Size of Reading Fixations," *Psychological Record*, 1937, 1, No. 28.

12) G. H. Fairbanks, "The Relation Between Eye Movements and Voice in the Oral Reading of Good and Poor Silent Readers," *Psychological Monographs*, 1937, 48, 78–107.

13) J. C. Flanagan, "A New Type of Reading Test for Secondary-School and College Students Which Provides Separate Scores for Speed of Comprehension and Level of Comprehension," *Official Report of the American Educational Research Association*, 1938, 195–199.

14) W. S. Foster, "The Effect of Practice on Visualizing and upon the Reproduction of Visual Impressions," *Journal of Educational Psychology*, 1911, 2, 11–22.

15) F. M. Garver and R. D. Matthews, "An Analysis of the Results of Speed Drills with the Metronoscope to Increase Reading Rate," *Journal of Educational Psychology*, 1939, 30, 693–698.

16) A. I. Gates, "Rôle of Visual Perception, Intelligence, and Associative Processes in Reading and Spelling," *Journal of Educational Psychology*, 1926, 17, 433–445.

17) M. D. Glock, "The Effect upon Eye Movements and Reading Rate at the College Level of Three Methods of Training," Ph. D. Thesis, State University of Iowa, Iowa City, 1947.

18) C. T. Gray, "Types of Reading Ability as Exhibited Through Tests and Laboratory Experiments," *Supplementary Educational Monographs*, Chicago: University of Chicago Press, 1917, 1, No. 5.

19) W. S. Gray, "Summary of Reading Investigations," *Journal of Educational Research*, 1936, 29, 407–432; 1937, 30, 553–576; 1938, 31, 401–434; 1939, 32, 481–517; 1940, 33, 481–523; 1941, 34, 401–443; 1942, 35, 401–442; 1943, 36, 401–444; 1944, 37, 401–440; 1945, 38, 401–429; 1946, 39, 401–453.

20) ———, *Reading*, in W. S. Monroe (Ed.) *Encyclopedia of Educational Research*, New York: The Macmillan Co., 1941, 891–926.

21) H. A. Greene, A. N. Jorgensen, and V. H. Kelley, *Iowa Silent Reading Tests (New Edition)*, Yonkers-on-the-Hudson; N. Y.: World Book Company, 1939.

22) R. G. Gundlach, D. A. Rothschild, and P. T. Young: "A Test and Analysis of 'Set,'" *Journal of Experimental Psychology*, 1927, 10, 247–280.

23) A. J. Harris, *How to Increase Reading Ability*, New York: Longmans, Green and Co., 1940.

24) H. A. Imus, J. W. M. Rothney, and R. M. Bear, *Evaluation of Visual Factors in Reading*, Hanover, N. H.: Dartmouth College Publications, 1938.

25) C. A. Knehr, "The Effects of Monocular Vision on Measures of Reading Efficiency and Perceptual Span," *Journal of Experimental Psychology*, 1941, 29, 133–154.

26) G. F. Kuder and M. W. Richardson, "The Theory of the Estimation of Test Reliability," *Psychometrika*, 1937, 2, No. 3, 151–160.

27) E. F. Lindquist, *Statistical Analysis in Educational Research*, Boston: Houghton Mifflin Co., 1940.

28) O. F. Litterer, "An Experimental Study of Visual Apprehension in Reading," *Journal of Applied Psychology*, 1933, 17, 266–276.

29) J. M. McCallister, *Remedial and Corrective Instruction in Reading*, New York: D. Appleton-Century Co., 1936.

30) M. C. Pollock and L. C. Pressey, "An Investigation of the Mechanical Habits in Reading of Good and Poor Readers," *Educational Research Bulletin*, 1925, 4, 273–275.

31) L. C. Pressey and M. Truax, "Value of Drill in the Mechanics of Reading," *Educational Research Bulletin*, 1928, 8, 310–312.

32) S. Renshaw, "The Visual Perception and Reproduction of Forms by Tachistoscopic Methods," *Journal of Psychology*, 1945, 20, 217–232.

33) Sister Mary Riley, (of the Visitation), "Visual Perception in Reading and Spelling: A Statistical Analysis," *Educational Research Bulletin*, Catholic University of America, 1929, 4, No. 1.

34) F. P. Robinson, "Role of Eye Movements in Reading, with Evaluation of Techniques for Their Improvement," State University of Iowa Studies, 1933, No. 9.

35) ———, "The Tachistoscope as a Measure of Reading Perception," *American Journal of Psychology*, 1934, 46, 132–135.

36) M. V. Rogers, "Comprehension in Oral and Silent Reading," *Journal of General Psychology*, 1937, 17, 394–397.

37) C. H. Seivers, "The University of Wichita Freshman Reading Program," University of Wichita Bulletins, 1939, 14, No. 6.

38) ———, "Study III: Wichita Public School Reading Program," University of Wichita Bulletins, 1939, 14, No. 7.

39) J. P. Seward, "The Effect of Practice on the Visual Perception of Form," *Archives of Psychology*, 1931, 20, No. 130.

40) E. W. Siebert, "Reading Reactions for Varied Types of Subject-matter; An Analytic Study of the Eye Movements of Eighth-grade Pupils," *Journal of Experimental Education*, 1943, 12, 37–44.

41) E. D. Sisson, "The Rôle of Habit in Eye Movements in Reading," *Psychological Record*, 1937, 1, 159–168.

42) ———, "Eye-Movement Training as a Means of Improving Reading Ability," *Journal of Educational Research*, 1938–9, 32, 35–41.

43) ———, "The Causes of Slow Reading: An Analysis," *Journal of Educational Psychology*, 1939, 30, 206–214.

44) L. G. Stone, "Reading Reactions for Various Types of Subject-matter: An Analytical Study of the Eye Movements of College Freshmen," *Journal of Experimental Education*, 1941–2, 10, 64–77.

45) R. Strang, "Diagnosis and Remediation," in Gray, W. S., *Reading in General Education*, Washington, D.C., American Council on Education, 1940, 337–338.

46) J. B. Stroud, "A Critical Note on Reading," *Psychological Bulletin*, 1942, 39, 173–178.

47) J. B. Stroud, and M. Henderson, "Rate of Reading and Learning by Reading," *Journal of Educational Psychology*, 1943, 34, 193–205.

48) J. B. Stroud, "Rate of Visual Perception as a Factor in Rate of Reading," *Journal of Educational Psychology*, 1945, 36, 487–498.

49) ———, *Psychology in Education*, New York, Longmans, Green and Co., 1946.

50) A. J. Sutherland, "The Relationship between Perceptual Span and Rate of Reading," *Journal of Educational Psychology*, 1946, 37, 373–380.

51) D. E. Swanson, "Common Elements in Silent and Oral Reading," *Psychological Monographs*, 1937, 48, No. 3, 36–60.

52) E. A. Taylor, *Controlled Reading: A Correlation of Diagnostic Teaching and Corrective Techniques.* Chicago: University of Chicago Press, 1937.

53) L. L. Thurstone, *Identical Forms Test, Research Form IF47*, Science Research Associates, 228 So. Wabash Ave., Chicago 4, Illinois, 1947. See also "Primary Mental Abilities," *Psychometric Monographs*, 1938, No. 1.

54) M. A. Tinker, "The Rôle of Eye Movements in Reading with an Evaluation of Techniques for Their Improvement, a Review," *Journal of Educational Psychology*, 1933, 24, 555–556.

55) ———, "Reliability and Validity of Eye-Movement Measures in Reading," *Journal of Experimental Psychology*, 1936, 19, 732–746.

56) ———, "Time Taken by Eye Movements in Reading," *Journal of Genetic Psychology*, 1936, 48, 468–471.

57) ———, "Motor Efficiency of the Eye as a Factor in Reading," *Journal of Educational Psychology*, 1938, 29, 167–174.

58) ———, "The Study of Eye Movements in Reading," *Psychological Bulletin*, 1946, 43, 93–120.

59) A. E. Traxler, "The Value of Controlled Reading; Summary of Opinion and Research," *Journal of Experimental Education*, 1943, 11, 280–292.

60) ———, "Ten Years of Research in Reading," *Educational Records Bureau Bulletin*, No. 32.

61) R. Y. Walker, "The Eye Movements of Good Readers," *Psychological Monographs*, 1933, 44, 95–117.

62) ———, "A Qualitative Study of the Eye Movements of Good Readers," *American Journal of Psychology*, 1938, 51, 472–481.

63) C. O. Weber, "The Acquisition and Retention of Reading Skills by College Freshmen," *Journal of Educational Psychology*, 1939, 30, 453–460.

64) ———, "The Effects of Practice on the Perceptual Span for Letters," *Journal of General Psychology*, 1942, 26, 347–351.

65) G. M. Whipple, "Effect of Practice upon the Range of Visual

Attention and of Visual Apprehension," *Journal of Educational Psychology*, 1910, 1, 249–262.

66) B. E. Witzeman, "An Experimental Study Using the Ophthalmograph and Metronoscope in the Diagnosis and Treatment of Reading Defects," *Journal of Psychology*, 1941, 11, 304–307.

67) R. S. Woodworth, *Experimental Psychology*, New York: Henry Holt and Co., 1938.

GENERAL, VERBAL, AND SPATIAL BI-FACTORS AFTER THREE YEARS

FRANCES SWINEFORD*
The University of Chicago

Longitudinal studies of the factors defined by methods of factor analysis should contribute significantly to our understanding of the nature of such factors. A report has been made of the results of testing one group of pupils after a one-year interval and a second group after a two-year interval.[1] Scores are now available for a third group who were given the same battery of nine tests in 1944 and again in 1947. At the first testing these eighty-eight pupils were in the sixth grade in seven different elementary schools. The second testing took place shortly after they had entered high school (ninth grade).

Three of the tests were selected because earlier research has shown them to measure only the general factor, which is common to all mental tests. These are (1) Arithmetic, (2) Series Completion, and (3) Deduction. Three more are correlated with both the general factor and the verbal factor. They are (4) General Information, (5) Paragraph Comprehension, and (6) Word Meaning. Finally, three tests measure the general and spatial factors: (7) Punched Holes, (8) Drawings, and (9) Visual Imagery.[2]

The intercorrelations of the nine tests (Table 1) have been analyzed for each testing. The same pattern plan is appropriate for both, but the ninth-grade factor coefficients are the higher in most instances (Table 2). These data will be employed to throw further light on such problems concerning growth in factor scores over a period of time as, first, whether any of the corresponding factors defined by the two analyses may be regarded as the same factor, and, second, what changes may have taken place in the three-year interval.

---

* Now with Educational Testing Service, Princeton, N. J.

[1] Frances Swineford, "Growth in the General and Verbal Bi-factors from Grade VII to Grade IX." *Journal of Educational Psychology*, 38, (May, 1947), 257–72.

[2] Descriptions of the tests are given in Frances Swineford, *A Study in Factor Analysis: The Nature of the General, Verbal, and Spatial Bi-factors.* Supplementary Educational Monographs, No. 67. Chicago: University of Chicago Press, 1948.

TABLE 1.—INTERCORRELATIONS OF NINE TESTS (N = 88)

Test	1	2	3	4	5	6	7	8
Grade VI								
General:								
1.....								
2.....	.452							
3.....	.444	.448						
Verbal:								
4.....	.406	.418	.315					
5.....	.488	.450	.228	.614				
6.....	.392	.469	.359	.648	.565			
Spatial:								
7.....	.372	.546	.482	.515	.415	.426		
8.....	.323	.411	.347	.416	.359	.438	.715	
9.....	.329	.498	.294	.372	.388	.339	.466	.409
Grade IX								
General:								
1.....								
2.....	.543							
3.....	.494	.635						
Verbal:								
4.....	.422	.580	.448					
5.....	.477	.640	.443	.769				
6.....	.390	.552	.444	.727	.717			
Spatial:								
7.....	.386	.645	.564	.491	.581	.505		
8.....	.373	.560	.459	.508	.543	.442	.764	
9.....	.358	.535	.478	.553	.582	.542	.806	.780

Two methods were employed in the earlier study to investigate the extent to which the factor estimates which were made at different times are comparable. One method is to treat the two patterns of Table 2 as though they were a single pattern with eighteen variables. With the exception of nine specific overlaps, one for each test, this pattern should reproduce, within sampling error, the correlations between the sixth-grade scores and the ninth-grade scores if the factors are actually the same. The observed correlation coefficients are given in Table 3. In order to bring together all the data of this sort, some of the results for

TABLE 2.—COMMON-FACTOR PORTIONS OF BI-FACTOR PATTERNS

Test	Grade VI			Grade IX		
	General	Verbal	Spatial	General	Verbal	Spatial
1.....	.634	....	....	.599		
2.....	.743	....	....	.850		
3.....	.572	....	....	.702		
4.....	.622	.575	....	.692	.554	
5.....	.601	.418	....	.758	.441	
6.....	.630	.445	....	.669	.477	
7.....	.723	....	.453	.743	....	.496
8.....	.591	....	.636	.669	....	.538
9.....	.580	....	.104	.693	....	.587

TABLE 3.—CORRELATIONS BETWEEN SIXTH-GRADE AND NINTH-GRADE SCORES (N = 88)

Grade IX	Grade VI								
	1	2	3	4	5	6	7	8	9
1.....	.703	.427	.437	.395	.367	.354	.430	.280	.323
2.....	.530	.602	.505	.615	.518	.547	.561	.468	.386
3.....	.499	.388	.623	.408	.376	.432	.559	.471	.420
4.....	.418	.363	.291	.693	.683	.596	.417	.397	.239
5.....	.444	.439	.363	.640	.724	.584	.456	.470	.226
6.....	.421	.505	.381	.666	.696	.708	.414	.354	.365
7.....	.349	.539	.364	.636	.532	.544	.786	.744	.423
8.....	.323	.462	.350	.497	.448	.432	.724	.807	.470
9.....	.342	.439	.378	.547	.528	.442	.777	.742	.493

the first two retested groups are presented in Table 4 for comparison with the current figures. The table includes the mean residual correlation and the standard deviation of the residuals. Three more tests are used in the present study than were available for the other two groups.

TABLE 4.—MEANS AND STANDARD DEVIATIONS OF "INTER-TESTINGS" RESIDUAL CORRELATIONS FOR THREE GROUPS

Statistic	Group A: Retested after One Year (212 cases)	Group B: Retested after Two Years (173 cases)	Present Group: Retested after Three Years (88 cases)
Number of residuals	30	30	72
Mean residual	−.0198	−.0165	−.0191
Standard deviation of residuals	.0440	.0780	.0718
Standard error of mean	.00817	.0145	.00852
M/σ_M	−2.42	−1.14	−2.24
$\sqrt{\Sigma \underline{r}^2/N}$	.0472	.0825	.0732

It is apparent from the mean residuals of Table 4 that the coefficients of Table 3 tend to be lower than probably would have been the case had there been no time interval between the two testings. The actual differences are extremely small—the means are all less than .02—but are nevertheless too consistent to be reasonably ascribable to chance alone. They probably reflect differences in rates of growth in the various areas tested. Since the means are not zero, a more valid test of the general appropriateness of the pattern than is afforded by the standard deviation of the residuals would be given by their variation about zero rather than about the observed means. The values $\sqrt{\Sigma \underline{r}^2/N}$ thus appear in the last line of the table, where $\underline{r}$ is a residual correlation, and N is the number of residuals summed. Even these measures, when compared with the standard error of a zero correlation, indicate that the patterns can be accepted as satisfactory descriptions of the 'inter-testings' correlations. Thus it has been shown that the factor patterns derived separately from the data of the two testings made three years apart yield the

same factors, since, set down as a single pattern for eighteen variables, they reproduce the inter-testings correlations extremely well with only minor disturbances probably due to variable rates of growth among the pupils.

The second method which was used in the earlier study to compare the factors obtained in the two testings was the calculation of the correlations between the estimates of the corresponding factors. Nine-variable regression equations were written in standard form for estimating each factor from the data of each testing, six equations in all. In theory the bi-factors are uncorrelated. Their estimates, however, have some intercorrelation because the predictions fall short of perfection. The correlations were computed by multiplying each pair of regression equations, summing over the eighty-eight cases, and dividing by eighty-eight times the product of the corresponding standard deviations. The resulting correlations are presented in Table 5.

TABLE 5.—INTERCORRELATIONS OF FACTOR ESTIMATES (MULTIPLE CORRELATIONS IN DIAGONAL)

Factor	Grade VI			Grade IX		
	$\overline{G}$	$\overline{V}$	$\overline{S}$	$\overline{G}$	$\overline{V}$	$\overline{S}$
Grade VI:						
$\overline{G}$.....	*.912*					
$\overline{V}$.....	.196	*.761*				
$\overline{S}$.....	.205	−.316	*.779*			
Grade IX:						
$\overline{G}$.....	.862	.243	.230	*.935*		
$\overline{V}$.....	.101	.637	−.277	.184	*.803*	
$\overline{S}$.....	.275	−.164	.697	.168	−.354	*.834*

The two general-factor estimates have a correlation of .86. Corresponding figures for the groups tested after one and two years, respectively, are .89 and .83. This evidence provides additional support for the hypothesis that the general factor remains the same throughout the period under consideration.

The verbal factor exhibits less stability. The three correlations are .78 and .74 for the earlier groups and .64 for the present one. The only correlation available for the spatial factor is that of .70 from Table 5. Lower values are to be expected for group factors than for the general factor for two reasons; namely, because their estimates are less reliable than general-factor estimates (the multiple correlations are lower), and because the pattern weights themselves are subject to larger sampling error. It should also be noted that the values off the diagonals are of the same order between the two testings as within either testing. Thus there are no adequate grounds for rejecting the hypothesis of constant group factors.

The second problem under consideration relates to the changes which have taken place during the three-year period. All the test means, of course, show statistically significant increases. In order to check factor estimates for increases it is necessary that the estimates for both testings be made from the same regression equation. The equations were first written in raw-score form and transmuted to yield means of 50 and standard deviations of 10 in each instance. The sixth-grade test means were then substituted in the ninth-grade equations, and *vice versa*. Substitutions were also made separately for the thirty-five girls and the fifty-three boys who constitute the total group. The results are listed in Table 6.

TABLE 6.—MEAN FACTOR ESTIMATES FOR GIRLS, BOYS, AND TOTAL GROUP FROM ALL REGRESSION EQUATIONS

Equation	Grade for Which Estimates Are Given						Increase		
	Grade VI			Grade IX					
	Girls	Boys	Total	Girls	Boys	Total	Girls	Boys	Total
Grade VI:									
$\bar{G}$.....	49.6	50.3	50.0	61.9	62.8	62.4	12.3	12.5	12.4
$\bar{V}$.....	48.5	51.0	50.0	52.4	54.7	53.8	3.9	3.7	3.8
$\bar{S}$.....	47.7	51.5	50.0	46.8	50.8	49.2	− 0.9	− 0.7	− 0.8
Grade IX:									
$\bar{G}$.....	42.3	42.7	42.5	49.5	50.3	50.0	7.2	7.6	7.5
$\bar{V}$.....	45.0	47.5	46.5	49.7	50.2	50.0	4.7	2.7	3.5
$\bar{S}$.....	45.4	47.7	46.8	47.1	51.9	50.0	1.7	4.2	3.2

Standard deviations have not been computed for these data. Previous work has shown, however, that older groups are consistently more variable than younger groups. For this reason the increases based on sixth-grade equations should be larger than those based on ninth-grade equations. It is obvious from Table 6 that the most important increase occurs for the general factor. This finding agrees with the results of a cross-section-type study made earlier.[1]

The verbal factor exhibits an increase for each group when either of the regression equations is used. The increase is smaller in all instances than that for the general factor. This finding, also, agrees with the earlier one, where the verbal factor showed a definite increase from a fifth-grade group to a tenth-grade group but not without minor discrepancies.

Curious results are obtained for the spatial factor. In the earlier study the figures for this factor suggested a slight tendency for the boys' means to increase with increasing maturity, but provided no consistent results for girls. Here, the results by the two equations are contradictory. When the sixth-grade equation is used, a decrease appears in mean score over the three-year period. When the ninth-grade equation is used, on the other hand, there is an increase comparable to that for the verbal factor. A major portion of the discrepancy is probably attributable to Test 9. This test proved quite difficult for sixth-grade pupils. The sixth-grade factor pattern (Table 2) shows it to be substantially related to the general factor, but its correlation with the spatial factor is very low and highly unstable. Indeed, its spatial-factor regression coefficient is a small negative, showing that this test contributes to the spatial-factor estimate nothing that is not already included in the scores on Tests 7 and 8.

By the time the pupils have reached the ninth grade, Test 9 is no longer too difficult for them. It now is related to the spatial factor to an appreciable degree. A low score would now indicate low spatial ability, whereas at the earlier level a low score indicated low general ability and was almost independent of spatial ability. It would appear, therefore, that, even assuming the factors measured at different levels to be the same, regression equations based on data obtained at one level may not be entirely

---

[1] *Ibid.*, ch. v.

appropriate for scores at another level of ability. In the present case, Test 9 seems to be the chief offender. If the spatial-factor equations are recomputed with Test 9 omitted, the increases of −.9, −.7, −.8, 1.7, 4.2, and 3.2 of Table 6 change to −.5, .1, −.1, 1.6, 1.2, and 1.4, respectively. Discrepancies still obtain, but they have been considerably reduced in magnitude. These results support the conclusion reached in the cross-section study; namely, that there is very little change apparent in the spatial factor during the period under consideration.

SUMMARY

Data for nine tests administered to eighty-eight pupils when they were in the sixth grade and again three years later have been analyzed in order to procure evidence relating to constancy and growth of three bi-factors. The three bi-factors are the general, verbal, and spatial factors, which are mutually uncorrelated. The analysis suggests the following generalizations:

1) The three bi-factors measured by the present test battery tend to remain the same factors at the ninth-grade level as they were at the sixth-grade level. Evidence supporting this conclusion derives from the readiness with which the two separate factor patterns may be combined to form a single pattern for eighteen variables having three common factors; and from the substantial correlation between the estimates of corresponding factors from the two separate analyses.

2) The three-year growth in factor scores, indicated by mean factor estimates, confirms the findings of an earlier cross-section study. In terms of the variability of the group, the general factor shows the greatest increase. The verbal factor, likewise, increases, but by a much smaller amount. No increase during this period is apparent for the spatial factor estimates.

3) Comparison of pattern weights and multiple-regression coefficients raises some question concerning the applicability of these tests at both levels. It is shown, for example, that one of the spatial tests is a much better predictor of S at the higher level than at the lower level. One may conclude, therefore, that regression equations based on data for a sample from one population are not necessarily equally applicable to scores taken from another population, even though the factors involved may be the same in both instances.

THE FUNCTION OF THE DIVISION OF EDUCATIONAL PSYCHOLOGY OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION: A COMMITTEE REPORT

VICTOR H. NOLL, Chairman
Michigan State College
JOHN E. HORROCKS
Ohio State University
and
G. LESTER ANDERSON
University of Minnesota

PURPOSE OF THE REPORT

At the 1948 meeting of the American Psychological Association the present committee was appointed by the Division of Educational Psychology for the dual purpose of considering the function of the Division, and its relation to the other divisions of the Association. In attempting to answer the two questions proposed to it, the Committee felt that it would first be necessary to consider the field of educational psychology as a whole as a means of gaining a better perspective against which to view the function and the relationships of the division. Hence, this report is presented in two parts. Part I attempts to view the broad field of educational psychology as a whole, and Part II attempts to define the function and responsibilities of the division in the light of the point of view developed in Part I.

PART I

WHAT EDUCATIONAL PSYCHOLOGY IS

Educational psychology takes its form and gets its meaning from the two fields of education and psychology. Education is a social process. Psychology is a discipline, a body of knowledge, a social and natural science.

Psychology has usually been defined as the study of behavior, generally in its molar rather than its molecular aspects. Behavior is studied to the end that it may be predicted and controlled.

Education as a social process is concerned with the establish-

ment of certain behavior patterns in men in order that they can adapt to a given environment, or in order that they may perform in certain ways so that the adaptation of other men is enhanced.

Educational psychology is concerned primarily with the study of human behavior as it is changed or directed under the social processes of education, and, secondarily, with those studies or processes that contribute to an increased understanding of how behavior is changed and directed through education.

Education is possible because humans can learn. Learning is a change in behavior which is correlated with experience. Education is concerned with learning as it is provided in a controlled environment. The educational psychologist is thus concerned primarily with learning. The psychology of human learning is the central core of educational psychology. Thus the function of the educational psychologist becomes specifically concerned with the study of human learning and functioning, to the end that they may be guided, directed and carried out more successfully and efficiently.

It is not to be supposed, however, that educational psychology may properly confine itself only to the study of human learning as such. It must be remembered that this study and concern will take the educational psychologist into any aspect of psychology or education which contributes to his understanding of learning. Thus, everything he does has meaning as it relates to learning, whether it be measurement and evaluation, guidance and counseling, studying the psychology of school subjects, promoting social development, or performing a training function within industry. These areas are concerned with human beings in order to predict their capacity to learn—and hence, to be educated; to prepare them to learn efficiently, to evaluate the learning which has been presumed to occur in order to re-educate the learner who is deficient; or to improve teaching procedures, or to study certain specialized aspects of learning such as learning to read, or learning how to do a job.

Educational psychology is also concerned with non-learned types of behavior, (those which are maturational in origin) for two reasons. First, learned behavior and behavior which is maturational are closely related. Maturation must take place before certain processes of education can be effective. Those who attempt to guide or direct learning without an understanding of the maturational processes cannot help but be ineffective.

Second, maturation normally progresses under a 'normal' environment. If, in a desire to produce certain kinds of learned behavior, the environment is too much disturbed, the normal growth processes will be disturbed and psychological harm may ensue.

THE EDUCATIONAL PSYCHOLOGIST AS A SPECIALIST

Because the guidance, direction, control, and evaluation of learning, and the understanding of the normal functions of the learner are complex matters, both at the levels of understanding and of study, the educational psychologist is often placed in a position of having to take a segment of the larger field as his special sphere of operation. Some choose to be primarily concerned with learning about and applying the methods of educational psychology to the child, or the adult; the brilliant, or the handicapped. Some choose to study learning which is school produced and so specialize in the psychology of reading or arithmetic, or the other school subjects. Still others are interested in segments of learned behavior such as personality or attitudes. Others, in their desire to promote adjustment interest themselves in clinical psychology, or in guidance or counseling. However, whether he operates as a general educational psychologist or whether he espouses one or more specialties, the educational psychologist generally approaches his work with an over-all point of view based upon a somewhat common body of knowledge, and perception of the common problems of his profession. Despite the great diversity of activities which occupy the time of the educational psychologist, a frame of reference for his activities is supplied by the limits set by the two fields, education and psychology. That is, he is interested in studying some aspect of human behavior as it may be changed under the conditions of learning.

EDUCATIONAL PSYCHOLOGY IN TERMS OF ITS CONTENT

Another approach to an examination of educational psychology as a field would be an examination of the general fields or areas of psychology that must be part of the content and experience background of every educational psychologist. There would appear to be five such major areas: human growth and development, learning, personality and adjustment, measurement and evaluation, and techniques and methods in educational

psychology. Looking at each area in turn, the more important subheadings under each division are as follows:

Fields or Areas of Educational Psychology

I. Human Growth and Development—
   a) Heredity and Environment
   b) General Growth and Development
   c) Social and Emotional Development
   d) Motivation, Drives—Basic Theory
   e) Intelligence, Aptitudes, Interests
   f) Individual Differences

II. Learning—
   a) General Nature of Learning
   b) Factors Influencing Learning
   c) Motivation—Devices in Teaching
   d) Skills
   e) Reasoning and Problem-Solving
   f) Attitudes
   g) Learning of Particular School Subjects
   h) Transfer of Training

III. Personality and Adjustment—
   a) Emotions
   b) Mental Hygiene of the Pupil
   c) Mental Health of the Teacher
   d) Exceptional Children
   e) Character
   f) Social Interaction

IV. Measurement and Evaluation—
   a) Basic Principles of Measurement
   b) Measurement of Intelligence and Aptitudes
   c) Measurement of Learning
   d) Measurement of Adjustment
   e) Application of Results of Measurement

V. Techniques and Method in Educational Psychology—
   a) The Scientific Study of Educational Problems
   b) Statistical Techniques
   c) Implementation of Research for the Classroom Teacher

EDUCATIONAL PSYCHOLOGY IN TERMS OF AREAS OF OPERATION

The definition of the field of educational psychology may also be approached from the standpoint of the jobs held by those who

are commonly designated as educational psychologists, and whose training has been particularly strong in those aspects of psychology designated previously as falling within the area of competence of the educational psychologist. An analysis would appear to indicate a number of types of activity, one or more of which usually fall within the duties of the person called an educational psychologist.

First, is the teaching function at the college level. Ordinarily courses taught by educational psychologists are designed for teachers in training or in service, for persons studying to become educational psychologists, for persons entering other areas of psychology, and for persons electing the course for cultural reasons or as part of their training in an allied field such as sociology. Many, but not all of these courses are directly oriented toward the problems encountered in public school teaching and an understanding of school-age boys and girls.

Second, is the research function. Nearly all educational psychologists actually do research in one aspect or another of their field. Some act as directors of research, while others take as their primary responsibility the integration and interpretation of existing research as it concerns or is of use in one or more of the various branches of educational psychology.

Third, is the function of a personnel worker or expert. This may take the form of a training function in the preparation of guidance workers, and counselors or school psychologists. It may consist of student personnel services on the college or university level as a counselor or a remedial aids expert. It may take the form of a psychometrician or college guidance worker dealing among other things with selection, prognosis, placement, or transfer. The educational psychologist may also find himself fulfilling the capacity of a clinician, particularly where children are concerned.

Fourth, is the consulting function. This may range from consultation on educational and psychological problems in industry or the armed forces to field work in the public schools.

Fifth, is the function of psychological services administration. This function may range from a dean's office, or head of a psychological services center, to the administration of various psychological service programs.

Sixth, is the function of test construction and the evaluation of measuring instruments.

THE EDUCATIONAL PSYCHOLOGIST FUNCTIONS AS A SCIENTIST AND AN INTEGRATOR

Implicit in the above discussion is the point of view that educational psychology is a science in its own right with a content as well as methods of procedure. The educational psychologist functions as a scientist among teachers of education and acts, both in psychology and education, as an intermediary between 'pure' research and its application. His preoccupation is with a science of education, building his procedures and his points of view upon verified facts and research. He has, in a real sense, an integrative function that can be far-reaching service.

As has already been indicated, the study of human behavior under conditions that interest the educational psychologist is complex. There is an undeniable need for specialization. But there is also an undeniable need for an integration of the specialized tools and techniques, principles and generalizations, knowledge and information that are produced by persons working at segments of the field.

An integration of education and psychology serves two purposes. It focuses the total outcomes upon the single objective of improved learning for more effective adaptation of human beings to the exigencies of life. Marshalled forces are more effective than are isolated forces. Educational psychology marshalls and integrates the forces of education and of psychology. Second, integration is mutually enriching to the integrated segments. It shows limitations or weaknesses and also strengths. It produces criteria for judgments of value or worth. It makes relative judgments possible. In this instance, it is mutually beneficial to education and psychology in obvious ways.

Unfortunately, at the present time, with clogged channels of communication between educational psychologists and educationists on the one hand, and among educational psychologists themselves on the other, it becomes increasingly difficult to fulfill the integrative or the facilitating function. The question as to how these channels may be opened is a vital one in the area of educational psychology—as vital to education, however, as to psychology because the histories of education and psychology abound with the contributions of educational psychology.

Part II

BASIC OBJECTIVES OF THE DIVISION

The over-all function of the Division of Educational Psychology is the promotion of the guidance and direction of human learning and functioning. To attain this goal, it is held that the Division should work toward the attainment of five basic objectives: definition of the field of educational psychology, advancement of the interests of educational psychology as a profession, coördination with other divisions and organizations interested in the fields of education and psychology, dissemination of information, and the promotion of research.

Part I of this report constitutes an effort on the part of the present committee to define the field of educational psychology. As far as the Division is concerned this function may be carried further by various types of activities since it is not to be inferred by the members that this report settles the question permanently. The field is changing constantly and the present report is probably best regarded as a point of departure for further study and clarification. To this end, the Division may wish to engage in a normative survey of the duties and responsibilities of educational psychologists. It may also wish to sponsor some symposia in connection with the annual meetings with representatives from other divisions to further define the area of educational psychology.

ADVANCEMENT OF THE INTERESTS OF EDUCATIONAL PSYCHOLOGY AS A PROFESSION

(a) It cannot be too strongly emphasized that the Division is in a strategic position to assume a real leadership function in the promotion of educational psychology and the things for which it stands. There are many people in psychology who think of themselves primarily as educational psychologists, and while the pursuit of a specialty may encourage other interests or professional affiliations, their first allegiance is to educational psychology as a broad field. It is believed that these people would welcome an active professional program by the Division of Educational Psychology, and would like to participate, especially as they felt that it was 'their' Division and their program.

Because the educational psychologist must function on such a broad front of training and knowledge, and because he occupies a position which constantly requires him to integrate the work and thinking of a number of cognate fields, including education and psychology, it is sometimes difficult to gain a clear and coherent picture of the whole professional area and its proper direction. It is felt that the Division may be of service in clarifying the areas and in forming a rallying center for those working in the field. For example, there are certain areas of evaluation and measurement such as the industrial, with which the educational psychologist may not be much concerned. However, he is concerned with evaluation and measurement in relation to the particular function of improving human learning. Therefore, this area of evaluation and measurement constitutes one of particular interest and responsibility for the educational psychologist. He has a right to preëmpt it, and the Division should promote it in various ways.

(b) It is particularly important that a vigorous policy of recruitment be pursued to the end that some of the more promising young persons entering the field of psychology may be encouraged to find their speciality in the field of educational psychology. It is felt that many who have entered other areas of specialization would have preferred educational psychology if it could have been presented to them as an ongoing, vigorously-organized field with strong leadership. This would be especially true if the challenge offered by the problems found in the field of educational psychology could be properly presented. It is vital that any profession look constantly to the future and the new men coming in are the ones who must carry the work forward.

(c) It is also felt that the Division has a responsibility to those who would employ educational psychologists and that some service might be made available to the end that employers could be made aware of what an educational psychologist is and what he should be able to do. It is realized that the Division has no jurisdiction over employers, but it might well study the problem of desirable standards and then act in an advisory capacity when requested. This is by no means a plea for certification under a body such as the American Board of Examiners, because it is felt that educational psychologists neither desire nor are ready for certification. It is a plea for consideration and discussion

by the Division of advisory standards of training and of practice in and for the profession.

COÖRDINATION WITH OTHER DIVISIONS OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION AND OTHER PROFESSIONAL ORGANIZATIONS

It is the function of the Division of Educational Psychology to open channels of communication with the other divisions of the American Psychological Association to the end that we may coördinate our activities and coöperate where coöperation is feasible and would appear to promote our respective interests and functions. The Division should offer leadership wherever possible and should continually be on the alert to find opportunities of pooling our resources with those of other groups working on problems of mutual interest. It should particularly concern itself with integrative areas that are being neglected by other more specialized divisions of the Association.

For example, the Division has an important function as a liason body between psychology and education and other related disciplines. It has been stated that educational psychologists believe that there is a science of education and that the best interests of education demand the application of the scientific method to its problems. The American Educational Research Association has similar interests and there may well be opportunities for joint action. Within the American Psychological Association itself there are a number of opportunities for joint action with other divisions. The Division of Educational Psychology might well act jointly with the Divisions of School Psychologists and of Guidance and Counseling in the setting up of desirable programs of training for school psychologists and guidance workers, as one example. Another would be joint consideration with the Division of Industrial Psychology of some of the problems posed by training programs in industry. That the Divisions on Measurement and Evaluation, on Childhood and Adolescence, and on Maturity and Old Age have interests in common with the Division of Educational Psychology is too obvious to require discussion in this report. Joint symposia such as the one held by the Division of Educational Psychology and the Division on the Teaching of Psychology at the 1948 convention of the Association offer rich opportunities for consideration of mutual problems.

THE DISSEMINATION OF INFORMATION

The Division has a responsibility to disseminate information to its membership and to other interested persons in or out of psychology. One of the major functions of the educational psychologist is that of interpreting what is known from other areas of psychology to the teacher, and of showing him how he can make applications of what is known in the improvement of his teaching and his handling of boys and girls. Another function is that of applying his knowledge and techniques to new research. Anything that the Division can do to pass along what is being done in educational psychology will be a real contribution. Certainly there should be a division newsletter, not only on problems and information of the division, but also on the field as a whole. This committee would like to emphasize strongly the desirability of divisional sponsorship of a professional journal such as the *Journal of Educational Psychology* and would urge strong representations in the proper quarters to that end. A planned series of articles in other educational and professional journals might also be sponsored by the division. The possibilities of joint symposia and coöperation with professional educational organizations have already been indicated.

THE PROMOTION OF RESEARCH

The promotion of research would appear to be one of the most important and significant functions that the Division can assume. A planned research program in the form of a series of research projects could well be sponsored by the Division. Especially pertinent would be an analysis of needed research in educational psychology and possibly a nation-wide coöperative attack upon certain designated problems. It should be possible to encourage research in small institutions by including interested members of their faculties on research teams. It is possible that well-planned programs might attract foundation funds. The Division could set up a clearing-house for research information, and might sponsor an annual review of research in designated areas of educational psychology. An annual list of theses and dissertations in educational psychology, completed or underway, would constitute a useful contribution to the literature of educational research.

AN APPLICATION OF CHI-SQUARE TO THE DETERMINATION OF THE DISCRIMINATING POWER OF TEST QUESTIONS[1]

ORRIS C. HERFINDAHL

University of Illinois

In the rather extensive literature on the problem of discrimination there seems to have been little attempt to use the tool provided in Chi-square. Chi-square can provide a solution to the problem that is simple in the sense that it keeps calculations to a bare minimum and gives results in a form which can be easily interpreted by the teacher who knows little statistics and who, if any item analysis is to be done at all, must do it himself. Direct calculation of Chi-square for each question would require a prohibitive amount of labor. The purpose of this paper is to develop a chart which will permit the use of Chi-square with a minimum of calculation.[2]

Discriminating power is defined as the power of a question to distinguish better from poorer students. If a question discriminates perfectly at a certain point in an achievement scale, the better students will all answer correctly and the poorer students will all answer incorrectly. If it has no discriminating power, the better students will answer incorrectly just as frequently (in a relative sense) as the poorer students.

Consider the procedure involved in analyzing the questions of an 'objective' examination that has been given to N students. Rank the students according to their scores on the whole examination or according to some other criterion of achievement as shown in the first column of Table I. Then tabulate the answers by rights and wrongs. For each question, R students will have answered correctly and W students incorrectly. If, for each question, a horizontal line is drawn between the Rth and

[1] The writer is grateful to Dorothy C. Adkins for very helpful comments.

[2] For an earlier and similar solution to this problem, see J. P. Guilford, "The Phi Coefficient and Chi Square as Indices of Item Validity," *Psychometrika*, Vol. 6 (1941), pp. 11–19. His abac suffers from the facts that the number of N's (size of sample) is too small and the probability levels for χ^2 are too few. There is also some lack of precision in the construction of the abac as evidenced on page 16.

$(R + 1)$th students, the answers will be tabulated in a double dichotomy, the dichotomy in one direction being better and poorer students and in the other direction right and wrong answers on this particular question. Designate the observed frequencies in the cells by r_1 (the number of better students answering correctly), w_1 (the number of better students answering incorrectly), r_2 (the number of poorer students answering correctly), w_2 (the number of poorer students answering incorrectly), and totals by R, W, and N.

TABLE I

Rank of Student According to Achievement Criterion	Question x: Right Answers	Question x: Wrong Answers	Total
1 2 : R	r_1	w_1	R
R + 1 : N	r_2	w_2	W
Total	R	W	N

If the question has no discriminatory power, $\frac{r_1}{R} = \frac{R}{N}$, and similarly for the other cells. That is to say, the better students are just as likely to answer the question incorrectly as are the poorer students. It is desired to determine whether observed departures from frequencies determined on the hypothesis of no discrimination could occur by chance. This can be done by the use of Chi-square.

Since, for a double dichotomy,

$$\chi^2 = \frac{N(r_1w_2 - w_1r_2)^2}{R^2W^2} = \frac{N\left(\frac{r_1w_2}{N \cdot N} - \frac{w_1r_2}{N \cdot N}\right)^2}{\frac{R^2}{N^2} \cdot \frac{W^2}{N^2}},$$

it is evident that it makes no difference whether absolute or relative frequencies are used. Because there is but one degree of freedom in a double dichotomy, this equation can be expressed in terms of the variables $\frac{w_1}{N}$ and $\frac{R}{N}$ with the result[1]

$$\frac{w_1}{N} = \frac{R}{N}\left(1 - \frac{R}{N}\right)\left[1 - \left(+\sqrt{\frac{\chi^2}{N}}\right)\right].$$

A graph can now be prepared with χ^2 held at certain probability levels and with N given various values. The graph will be a series of second degree parabolas, and the hypothesis of no discrimination will also be represented by a parabola,[2]

$$\frac{w_1}{N} = \frac{R}{N}\left(1 - \frac{R}{N}\right).$$

At this point it may be noted that the negative root of $\frac{\chi^2}{N}$ will give parabolas in the region of negative discrimination, but this region is of no interest for the present problem.

For each question, the following steps are involved in the use of this procedure:

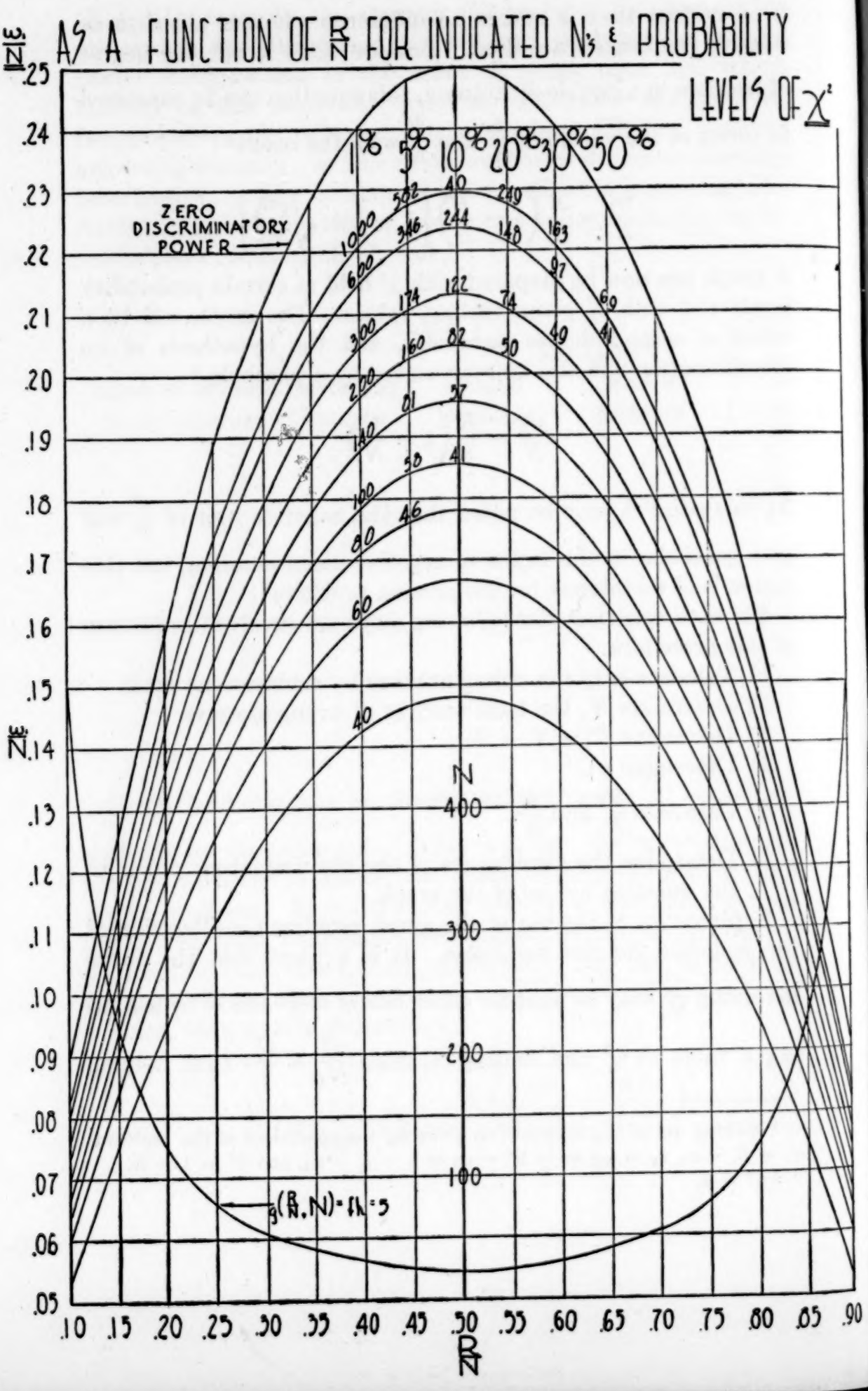
1) Tabulate only the wrong answers by achievement rank.
2) Determine W, the total number of wrong answers.
3) Determine $R(= N - W)$.
4) Determine w_1.
5) Calculate $\frac{w_1}{N}$ and $\frac{R}{N}$.
6) Determine the significance of the discriminatory power of the question by use of the graph.

Explanation of the use of the graph requires consideration of the equation for the parabolas. It is evident that the curve involving $\frac{\chi_1^2}{N_1}$ may be used for other values of χ^2 and N by assigning a value to χ_j^2 and finding N_j from $\frac{\chi_1^2}{N_1} = \frac{\chi_j^2}{N_j}$, thus holding

---

[1] Making use of the information given by the conditions of the problem: $r_1 = R - w_1$, $r_2 = w_1$, $w_2 = W - w_1 = 1 - R - w_1$, and $W = 1 - R$.

[2] $\chi^2 = 0$.


AS A FUNCTION OF R/N FOR INDICATED N's & PROBABILITY LEVELS OF χ²
1% 5% 10% 20% 30% 50%
ZERO DISCRIMINATORY POWER
1000
600
300
200
140
100
80
60
40
382
346
244
174
160
122
82
81
50
50
41
46
40
240
148
74
50
37
41
163
97
69
49
41
N
400
300
200
100
$\chi^2_3(R/N, N) = 11.345$
SIN
.25
.24
.23
.22
.21
.20
.19
.18
.17
.16
.15
.14
.13
.12
.11
.10
.09
.08
.07
.06
.05
.10 .15 .20 .25 .30 .35 .40 .45 .50 .55 .60 .65 .70 .75 .80 .85 .90
R/N

constant the member of the above equation which is under the radical although permitting χ^2 and N to vary. To use the graph consider a question that is represented by the point, $\left(\frac{R}{N}, \frac{w_1}{N}\right)$. This point is between two parabolas. Keeping the same relative distance between these curves, move the point visually until it is above the highest N on the chart (say 116) that is smaller than the actual N for this test. Then interpolate between the 116 and the N directly above it (say 174) for the actual N. All four of these N's are now at a certain probability level for χ^2. If the transferred point is below the interpolated point, χ^2 for this column's probability level is the greatest (of those considered on the chart) that is exceeded by χ^2 for this question. If, however, the transferred point is above the interpolated point, move the point to the next column to the right and repeat the process.

Consider two specific examples:

1) $N = 275$
$\frac{R}{N} = .67$ and $\frac{w_1}{N} = .19$

Move the point to the left, keeping the same relative distance between the two parabolas, until it is above 174, the highest N that is below the actual N of 275. Interpolate between 174 and 346 for 275 (linear interpolation will be accurate enough for the present purpose). Since the transferred point is below the interpolated point, the 5 per cent value of χ^2 is the highest (of those considered on the chart) that is exceeded by χ^2 for this question.

2) $N = 350$
$\frac{R}{N} = .65$ and $\frac{w_1}{N} = .20$

Move the point until it is above 300. Interpolate for 350. Since the transferred point is above the interpolated point, χ^2 for this question does not exceed the 1 per cent value of χ^2.

Now move the point until it is above 174. The transferred point is now below the interpolated point for 350, hence the 5 per cent value of χ^2 is the highest that is exceeded by χ^2 for this question.

χ^2 is not suitable if the hypothetical frequency in any cell is below, say, five. In the center of the chart is a scale for N

applying to a curve showing N as a function of the minimum and maximum R/N for which the lowest hypothetical frequency in one of the cells is five.[1] If N is 500, for example, the lowest R/N for which the test is suitable is .10 and the highest R/N is .90.

If only questions are selected which have a χ^2 exceeding, say, the 5 per cent value, results will be the same as with any statistical test. That is to say, approximately 5 per cent of the accepted questions will be, in fact, non-discriminatory.

Chi-square, as used here, shares a disadvantage of all dichotomous tests in that it does not completely reveal the relation between item validity and achievement level since intra-cell relationships are neglected. At the same time, it requires no assumption with respect to the form of the distribution. The dividing line on the achievement scale is not at an arbitrary point but is at a point with a rational basis; namely, that point indicated by the null hypothesis.[2] Perhaps its most important advantage, however, is that it gives results in a probability form which the statistically unsophisticated can interpret in a useful manner even though the interpretation may be incorrect from a theoretical point of view. The influence of size of sample on the utility of the results is taken account of, and this is of special importance for the individual teacher who will not have large samples at his disposal.

In any application of statistical method, a fundamental distinction is that between the probability of association and the degree of association between two variables. In particular, it

---

[1] The effective bounds are:

$$\text{a) } N = \frac{5}{(R/N - 1)^2}$$

$$\text{b) } N = \frac{5}{(R/N)^2}$$

[2] The desirability of this dividing point as against, say, the median total score, has been pointed out by Joseph Zubin. On general grounds and by the use of special illustrative cases he showed that if the median is used as the dividing point, misleadingly low values of χ^2 will be secured. At the same time, if the dividing point for the criterion is given by the null hypothesis rather than the median total score, he rejected Chi-square as a practical means of item validation because of the computational burden. See Joseph Zubin, "The Method of Internal Consistency for Selecting Test Items," *Journal of Educational Psychology*, vol. 25 (1934), pp. 352–6.

has often been said that Chi-square provides a measure of only the probability of association and not the degree of association as would be provided by, for example, the bi-serial r.[1] As a general statement, this is quite acceptable. It should be pointed out, however, that for samples of a 'constant' size, Chi-square will vary directly with any measure of the degree of association, apart from intra-cell differences among questions along the criterion scale. Any measure of the degree of association will depend on the extent of departure from null hypothesis frequencies, and so also does Chi-square for samples of the same size.

For samples of varying sizes, the difficulty can not be escaped. A simple measure of the 'degree' of association, utilizing computations already made for the determination of Chi-square, would be the vertical distance between the parabola for zero discrimination and the point, $\left(\frac{R}{N}, \frac{w_1}{N}\right)$, for question x. But since the significance of this vertical distance depends on the magnitude of R/N, the distance must be divided by the height of the zero discrimination parabola. That is,

$$\text{Degree of association} = \frac{\left(1 - \frac{R}{N}\right)\frac{R}{N} - \frac{w_1}{N}}{\left(1 - \frac{R}{N}\right)\frac{R}{N}}$$

The limits of this measure are plus one for perfect discrimination $\left(\frac{w_1}{N} = 0\right)$ and zero for no discriminatory power $\left(\frac{w_1}{N} = \left(1 - \frac{R}{N}\right)\frac{R}{N}\right)$, neglecting negative discrimination.

[1] See, e.g., *ibid.*, n. 2, p. 345 and Guilford, *op. cit.*, p. 16.

THE CONSTANCY OF ITEM-TEST CORRELATION COEFFICIENTS COMPUTED FROM UPPER AND LOWER GROUPS*

JEROME E. DOPPELT and EDITH M. POTTS
The Psychological Corporation

In the construction of a test we always arrive at a point where some form of item analysis must be undertaken. Although there are numerous methods available for estimating the discriminating power of items, the one generally regarded with greatest favor is the determination of biserial coefficients of correlation between the items and the criterion. Guilford[2] has noted that there is a distinct tendency to use the biserial coefficients as the standard against which to check all other devices for item analysis. Occasionally, an outside criterion is used for validation of items, but more often the items are correlated with the total test scores, and the test-constructor uses these item-test biserial coefficients in making his judgments concerning items to be retained in the final form of the test.

After the biserial coefficients have been obtained, the test-constructor may bask happily in the thought that he has used all of his data to determine a very reputable statistic for which a formula is available giving the approximate standard error. But the determination of biserial coefficients is an expensive and time-consuming procedure. Most test-constructors are very willing to consider other methods of item analysis provided, of course, that there is reasonable assurance there will be no serious sacrifice of accuracy.

The biserial coefficient or the point biserial coefficient is actually a Pearson product-moment coefficient under special conditions. Another method which also estimates a product-moment correlation under special conditions is the determination of a coefficient from the upper and lower portions of the distribution. The studies of T. L. Kelley[3] have shown that for maximum discrimination the upper and lower portions should include twenty-seven per cent of the total group. Further work by J. C. Flanagan[1] has resulted in a table which permits the direct reading of the correlation coefficient when the table is

* Presented as a paper before the American Psychological Association, September 11, 1948.

entered with the proportions of successes in each of the upper and lower twenty-seven per cent. There is a great saving of time and cost when this method is used. The saving is especially pronounced when IBM answer sheets have been used for the test. The graphic item counter of the IBM test scoring machine will accumulate and record the number of correct responses to as many as ninety items in one run. The counts are printed on a record sheet. Record sheets may be obtained for the upper and lower groups, the frequencies converted into percentages, and the coefficients determined from Flanagan's table. The method is a considerable time-saver, but has a serious limitation. No formula is available for estimating the error of a coefficient thus obtained. This paper reports an attempt to determine by empirical means an indication of the error of an r obtained in this manner.

It is, of course, well known that the standard error is dependent upon the number of cases, and should N become sufficiently large, the standard error of any statistic diminishes to an insignificant value. The present study was done on samples of three hundred seventy cases, an N not so large that the standard errors of correlation coefficients may be expected to vanish and yet of a size often used in item analyses. In the tryout of the experimental form of a test it is not impractical to collect at least three hundred seventy cases. The number three hundred seventy was used in this study because twenty-seven per cent of it equals 100, which is the maximum number of cases that may be conveniently recorded on a single record sheet for the graphic item counter and, of course, the frequency for each item is also the percentage of successes in the group.

An experimental form of the Potts-Bennett General Information Test was used in the investigation. In 1947, this test was given to large numbers of female applicants for admission to schools of nursing in cities of Massachusetts, Connecticut, and New York. The subjects were, with few exceptions, at least high-school graduates with the great proportion of them drawn from the upper half of their respective high-school classes. The test consisted of one hundred fifty multiple-choice items, the answers to which were recorded on IBM answer sheets. As the tests were returned to The Psychological Corporation, three successive sub-samples of three hundred seventy cases each were taken from the first eleven hundred ten tests. The three sub-

samples were equivalent with respect to means and standard deviations: the mean scores were 94, 92, and 93 and the respective standard deviations were 20, 20, and 23. For each sub-sample, the upper and lower twenty-seven per cent were determined and item-test coefficients were obtained for every item from Flanagan's table. There were thus three coefficients for each item. In order to obtain a series of item-test coefficients which might be considered more stable measures and used as criterion values, the entire sample of eleven hundred ten cases, with three hundred cases in each of the upper and lower groups, was analyzed and a fourth set of item-test coefficients was determined.

Items were grouped in classes according to the item-test coefficients for the total sample of eleven hundred ten cases. Within each group there was included the set of three coefficients for each item (one from each sub-sample). The sub-sample coefficients were thus grouped on the basis of the criterion values for the items. Table I presents the frequency, mean and standard deviation of sub-sample coefficients in each criterion group. The mean of the sub-sample coefficients in each class was essentially the same as the mean of the criterion coefficients of the class. The standard deviations ranged from .059 to .075. These standard deviations are actually slight overestimates. Since every class was based on a range of five criterion coefficients, there would be some variance in the class even if the three sub-sample coefficients of each item had been identical with the corresponding criterion value. An estimate was made of the standard deviations under such ideal conditions. These were found to fall between .011 and .018. Actually, then, the obtained standard deviations might be considered as ranging from .05 to approximately .07. These were compared with the approximate standard errors of biserial coefficients for the average r of each class and based on three hundred seventy cases, the size of one complete sub-sample. The biserial standard errors range from .041 to .061. The standard errors of product-moment coefficients under similar conditions range from .035 to .048.

It appears that the item-test coefficients computed from the upper and lower twenty-seven per cent of a distribution of three hundred seventy cases have standard errors only slightly larger than those of biserial coefficients. Of particular interest is the fact that the band of error which must be thought of as sur-

rounding any obtained coefficient is not so wide as to make interpretation of the coefficient hazardous. The use of three hundred seventy or more cases for an item analysis by the upper and lower twenty-seven per cent method seems to yield coefficients which are sufficiently stable for most item analysis purposes.

TABLE I.—MEANS AND STANDARD DEVIATIONS OF SUB-SAMPLE COEFFICIENTS GROUPED ACCORDING TO CRITERION COEFFICIENTS OF CORRESPONDING ITEMS

Group	Range of Criterion r	Mean Criterion r	Sub-Sample Coefficients		
			N	Mean	SD
1	.25–.29	.267	42	.268	.075
2	.30–.34	.325	39	.329	.069
3	.35–.39	.375	39	.374	.059
4	.40–.44	.423	75	.424	.069
5	.45–.49	.474	66	.477	.064
6	.50–.54	.511	48	.511	.068
7	.55–.59	.574	51	.579	.072

It is obvious, of course, that generalizations from these data must be very guarded. There were only three samples and the frequency of each class was relatively small. The criterion coefficients, although they were based on a large number of cases, were determined from a group which included the samples. Yet the evidence seems to indicate that when we have a group of at least three hundred seventy cases, a number not unreasonably large for item analysis, the coefficients computed by the upper-lower twenty-seven per cent method are, for practical purposes, as useful as the more laboriously-computed biserial coefficients.

REFERENCES

1) J. C. Flanagan. "Item analysis by test scoring machine graphic item counter." *Proceedings of educ. res. Forum*, 1940, IBM Corporation.

2) J. P. Guilford. *Psychometric Methods.* New York: McGraw-Hill Book Co., 1936.

3) T. L. KELLEY. "Selection of upper and lower groups for the validation of test items." *J. educ. Psychol.*, 1939, 30, 17–24.

BOOK REVIEW

NORMAN CAMERON. *The Psychology of Behavior Disorders. A Biosocial Interpretation.* Boston: Houghton Mifflin Company, 1947, pp. 622.

In some respects, Dr. Cameron's volume on the behavior of social disorders is one of the most significant books of the year in the field of psychology. One is reminded of Murphy's *Experimental Social Psychology* and Gordon Allport's *"Personality"* by comparison. The common characteristic of these volumes is their resourcefulness. They represent systematized consistent approaches to synthesizing the knowledge which was being accumulated and needed interpretation for use. Cameron's book is characterized by its comprehensive and resourceful treatment rather than for remarkably original or revolutionary ideas. Students of John Dewey will readily take to this volume and will recognize in it a thorough-going illustrative treatment of Dewey's concept of unity of personality conceived in terms of social behaviorism. This volume can be considered a significant achievement in many respects. It does away completely with the mind-body dualism and yet it includes within the volume all that is significant in abnormal behavior. A welcome exclusion is a long section on symptomatology. Case history material included here is not only well chosen and illustrates the facts described, but it is selected relevantly and pointedly. The emphasis in the volume is upon established facts. Psychoanalytic concepts are not found here, but some of the psychoanalytic findings are made use of, though re-interpreted into biosocial terms.

The first six of the eighteen chapters can be considered as descriptive of a frame of reference. They include chapters on personality development; behavior organization and behavior pathology; language, thought and rôle-taking in behavior disorders; need, frustration, conflict; and basic adjustive techniques. The chapter on personality development is done compactly and well, and interpretations and suggestions included are given in a very humanized manner.

The emphasis in Cameron's orientation as well as in his treatment of all the syndromes throughout the book is on learning. The emphasis is also on human communication, described by

Cameron as ranging from exclusion to inclusion, over-exclusion being illustrated by the hysterical syndrome and over-inclusion by schizophrenia. Thinking is considered in terms of its biosocial nature. The consideration of rôle-taking in behavior disorders is excellent for purposes of understanding the general adjustments of man in his world in relationship to self-approval. Adjustive techniques considered include direct aggression and temper tantrums, simple withdrawal, fear and anxiety; such defense techniques as attention-getting, identification, compensation, rationalization, and projection; and such escape techniques as insulation, negativism, regression, repression and fantasy.

Syndromes or clusters of related signs or symptoms, with a chapter devoted to each of eight selected are considered by the author. These are hypochondriacal disorders, fatigue syndromes, anxiety disorders, compulsive disorders, hysterical inactivation and hysterical autonomy, paranoid disorders, schizophrenic disorders, manic and depressive disorders. There is also a chapter on reactions of cerebral incompetence which is considered as essentially non-psychiatric, unless the cerebrally-incompetent develops a behavior disorder. All of these syndromes and disorders are considered in biosocial terms both in diagnosis and treatment. The last chapter is on therapy as biosocial behavior. It includes a consideration of aims of office as well as hospital treatment, and a consideration of the therapist, the patient and the results in the therapeutic situation.

Psychoanalysts or the psychoanalytically-minded will miss here Freudian concepts or some detailed contributions of psychoanalytic institute workers, but the meaningful content of problems involved are discussed in spite of these omissions. The Edipus complex, for example, does not show up at all, but the facts of attitudes in development are well considered. Particularly well done is the chapter on rôle-taking. Throughout, Cameron's reaction is consistent. He is critical of metaphysical concepts, obtuse thinking, extremely critical, for example, of psychoanalytic concepts in the exaggerated implication of the disciples and is forever searching for sound established procedures for keeping his feet on the ground as he works his way through problems of behavior disorders. Emphasis is typically on how the behavior disorders are acquired; hence, much detailed con-

sideration to the learning process in all disorders. Generally it might be said that he does in consideration of behavior disorders, what James Harvey Robinson did in history; namely, search for how a person can get that way. If one searched for critical responses, it might be said that there is relatively little attention given here to alcoholism, delinquency, and criminality and though the biosocial interpretation emphasis is consistently treated, one is surprised to find Healy's contributions on delinquency or more recent echological studies of behavior practically neglected. This kind of omission can well be excused considering the wealth of material well covered in this book with a distinctly consistent frame of reference. The comprehensive inclusion of the facts of abnormal psychology interpreted in a manner that is intelligible, understandable, scientific and makes possible research that promises to unite people interested in behavior disorders by the application of scientific methods is an achievement. Dr. Carmichael's flattering introduction which concludes with a prediction that the book will certainly have a far-reaching and important effect upon the professional workers who are called upon to understand and deal with behavior disorders is warranted, all things considered.

H. MELTZER
Psychological Service Center
St. Louis, Missouri